I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to:

Assistant Commissioner for Patents
Washington, D.C. 20231

On February 11, 2003

By Diana M. Schaller
Diana M. Schaller

Attorney Docket No. 0217us210

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Jens Sigurd Okkels, *et al.*

Application No.: 09/896,896

Filed: June 29, 2001

For: Peptide Extended Glycosylated Polypeptides

Examiner: S. Liu

Art Unit: 1653

TRANSMITTAL OF CERTIFIED COPIES OF PRIORITY APPLICATIONS

Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

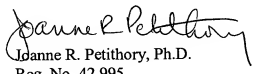
Applicants respectfully submit four Certified Copies of the following Priority Applications in the above-referenced application. All four claims of foreign priority were made in the original application papers in the above-referenced application.

COUNTRY:	APPLICATION NO.:	FILING DATE:
Denmark	PA 2000 01027	June 30, 2000
Denmark	PA 2000 01092	July 14, 2000
PCT	PCT/DK01/00090	February 9, 2001
PCT	PCT/DK00/00743	December 29, 2000

Applicants believe that no fee is due as the Certified Copies of the Priority Applications are being submitted prior to the payment of the issue fee. However, if the

Commissioner believes that a fee is due, the Commissioner is hereby authorized to charge the Processing Fee of $130.00, as set forth in 37 C.F.R. 1.17(i), to Deposit Account No. 50-0990. Please charge any additional fees associated with this paper, or credit any overpayment, to the above-noted Deposit Account.

Respectfully submitted,



Joanne R. Petithory, Ph.D.
Reg. No. 42,995

Maxygen, Inc.
Patent Department
515 Galveston Drive
Redwood City, California 94063
Telephone: (650) 298-5300
Facsimile: (650) 298-5446
Customer No. 30560

09/896,896

Kongeriget Danmark

Patent application No.: PA 2000 01027

Date of filing: 30 June 2000

Applicant: Maxygen ApS
Agern Allé 1
DK-2970 Hørsholm

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following document:

- The specification, claims and figures as filed with the application on the filing date indicated above.

Applicant has changed its name from ProFound Pharma A/S to Maxygen ApS with effect from 01 September 2000. A transcript from the Danish Companies Register has been submitted to the Danish Patent Office as proof of the change of name. Consequently, Maxygen ApS is the same Applicant as ProFound Pharma A/S.

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

Taastrup, 26 September 2002

Karin Schlichting
Head Clerk

PATENT- OG VAREMÆRKESTYRELSEN

FIELD OF THE INVENTION

The present invention relates to novel conjugates of an expanded polypeptide and a non-peptide moiety as well as means and methods for their preparation.

BACKGROUND OF THE INVENTION

Polypeptides, including proteins, are used for a wide range of applications, including industrial uses and human or veterinary therapy.

One generally recognized drawback associated with polypeptides is that they do not have a sufficiently high stability, are immunogenic or allergenic, have a reduced serum half-life, is susceptible to clearance, are susceptible to proteolytic degradation, and the like.

One method for improving properties of polypeptides has been to attach non-peptide moieties to the polypeptide to improve properties thereof. For instance, polymer molecules such as PEG has been used for reducing immunogenicity and/or increasing serum half-life of therapeutic polypeptides and for reducing allergenicity of industrial enzymes.

Glycosylation has been suggested as another convenient route for improving properties of polypeptides such as stability, half-life, etc.

Machamer and Rose, J. Biol. Chem., 1988, 263, 5948-5954 and 5955-5960, disclose modified glycoprotein G of viscular stomatitis virus that is glycosylated at additional N-glycosylation sites introduced in the polypeptide backbone.

US 5,218,092 discloses physiologically active polypeptides with at least one new or additional carbohydrate attached thereto. The additional carbohydrate molecule(s) is/are provided by adding one or more additional N-glycosylation sites to the polypeptide backbone, and expressing the polypeptide in a glycosylating host cell. Specifically modified G-CSF and urokinase molecules are disclosed.

US 5,041,376 discloses a method of identifying or shielding epitopes of a transportable protein, in which method an N-glycosylation site is introduced on the exposed surface of the protein backbone (using oligonucleotide-directed mutagenesis of the nucleotide sequence encoding the protein), the resulting protein is expressed, glycosylated and assayed for protein activity and for shielded epitopes.

WO 00/26354 discloses a method of reducing the allergenicity of proteins by including an additional glycosylation site in the protein backbone and glycosylating the resulting protein variant.

Guan et al., Cell, 1985, Vol. 42, 489-496 disclose glycosylated fusion protein variants comprising a rat growth hormone backbone C-terminally extended with transmembrane and cytoplasmic domains of the vesicular stomatitis virus glycoprotein, which growth hormone backbone has been modified to incorporate two additional N-glycosylation sites.

WO 97/04079 discloses lipolytic enzymes modified to by an N- or C-terminal peptide extension capable of conferring improved performance, in particular wash performance to the enzyme.

Matsuura et al., Nature Biotechnology, 1999, Vol. 17, 58-61 disclose the use of random elongation mutagenesis for improving thermostability of a non-glycosylated microbial catalase. The random elongation mutagenesis is conducted in the C-terminal end of the catalase.

In none of the above reference it has been disclosed or indicated that additional attachment groups for a non-peptide moiety of interest can be provided by an N- or C-terminal peptide extension so as to result in the preparation of polypeptide conjugates having improved properties. The present invention is based on this finding.

BRIEF DESCRIPTION OF THE INVENTION

Accordingly, in a first aspect the invention relates to a conjugate of a polypeptide and a non-peptide moiety, wherein the polypeptide part of the conjugate comprises the primary structure,

NH_2-X-P-COOH or NH_2-P-X-COOH,

wherein

X is a peptide addition comprising or contributing to an attachment group for the non-peptide moiety, and P is a polypeptide of interest.

In a second aspect the invention relates to a conjugate of a polypeptide and a non-peptide moiety, wherein the polypeptide part of the conjugate comprises the primary structure NH_2-P_x-X-P_y-COOH, wherein

P_x is an N-terminal part of a polypeptide P of interest,
P_y is a C-terminal part of said polypeptide P, and
X is a peptide addition comprising or contributing to an attachment group for the non-peptide moiety.

In other aspects the invention relates to a nucleotide sequence encoding the polypeptide part of a conjugate of the invention, an expression vector comprising said nucleotide sequence and methods of preparing a conjugate of the invention.

In a further aspect the invention relates to a method of improving (a) selected property/ies of a polypeptide P of interest, which method comprises a) preparing a nucleotide sequence encoding a polypeptide comprising the primary structure
NH_2-X-P-COOH or NH_2-P-X-COOH,
wherein
X is a peptide addition comprising or contributing to an attachment group for a non-peptide moiety that is capable of conferring the selected improved property/ies to the polypeptide P, when conjugated thereto,
b) expressing the nucleotide sequence of a) in an suitable host cell,
c) conjugating the expressed polypeptide of b) to the non-peptide moiety, and
d) recovering the conjugate resulting from step c).

DETAILED DISCLOSURE OF THE INVENTION

DEFINITIONS

In the context of the present application and invention the following definitions apply:

The term "conjugate" (or interchangeably "conjugated polypeptide") is intended to indicate a composite or chimeric molecule formed by the covalent attachment of one or more polypeptide(s) to one or more non-peptide moieties. The term covalent attachment means that the polypeptide and the non-peptide moiety are either directly covalently joined to one another, or else are indirectly covalently joined to one another through an intervening moiety or moieties, such as a bridge, spacer, or linkage moiety or moieties. Preferably, the conjugated polypeptide is soluble at relevant concentrations and conditions, i.e. soluble in physiological fluids such as blood. Examples of conjugated polypeptides of the invention include glycosylated polypeptides and PEGylated

polypeptides. The term "non-conjugated polypeptide" can be used about the polypeptide part of the conjugate.

The term "non-peptide moiety" is intended to indicate a molecule, different from a peptide polymer composed of amino acid monomers and linked together by peptide bonds, which molecule is capable of conjugating to an attachment group of the polypeptide of the invention. Preferred examples of such molecule include polymers, e.g. polyalkylene oxide or oligosaccharide moieties lipophilic groups, e.g. fatty acids and ceramides. The term "polymer molecule" is defined as a molecule formed by covalent linkage of two or more monomers and may be used interchangeably with "polymeric group". Except where the number of polymeric groups in the conjugate is expressly indicated, every reference to "polymer", "polymeric group" or "polymer molecule" referred to herein is intended as a reference to one or more polymeric groups of the conjugate.

The term "oligosaccharide moiety" is intended to indicate a carbohydrate-containing molecule comprising one or more monosaccharide residues, capable of being attached to the polypeptide (to produce a polypeptide conjugate in the form of a glycosylated polypeptide) by way of *in vivo* or *in vitro* glycosylation. The term "*in vivo* glycosylation" is intended to mean any attachment of an oligosaccharide moiety occurring *in vivo*, i.e. during posttranslational processing in a glycosylating cell used for expression of the polypeptide, e.g. by way of N-linked and O-linked glycosylation. Usually, the N-glycosylated oligosaccharide moiety has a common basic core structure composed of five monosaccharide residues, namely two N-acetylglucosamine residues and three mannose residues. The exact oligosaccharide structure depends, to a large extent, on the glycosylating organism in question and on the specific polypeptide. Depending on the host cell in question the glycosylation is classified as a high mannose type, a complex type or a hybrid type. The term "*in vitro* glycosylation" is intended to refer to a synthetic glycosylation performed *in vitro*, normally involving covalently linking an oligosaccharide moiety to an attachment group of a polypeptide, optionally using a cross-linking agent. *In vivo* and *in vitro* glycosylation are discussed in detail further below.

An "N-glycosylation site" has the sequence N-X'-S/T/C-X'', wherein X' is any amino acid residue except proline, X'' any amino acid residue that may or may not be identical to X' and preferably is different from proline, N asparagine and S/T/C either serine, threonine or cysteine, preferably serine or threonine, and most preferably threonine. An "O-glycosylation site" is the OH-group of a serine or threonine residue.

The term "peptide addition" is intended to indicate one or more consecutive amino acid residues that are added to the amino acid sequence of the polypeptide P of interest. Normally, the peptide addition is linked to the amino acid sequence of the polypeptide P by a peptide linkage.

The term "attachment group" is intended to indicate a functional group of the polypeptide, in particular of an amino acid residue thereof or an oligosaccharide moiety attached to the polypeptide, capable of attaching a non-peptide moiety of interest. Useful attachment groups and their matching non-peptide moieties are apparent from the table below.

Attachment group	Amino acid	Examples of non-peptide moiety	Conjugation method/Activated PEG	Reference
$-NH_2$	N-terminal, Lys	Polymer, e.g. PEG, with amide or imine group Lipophilic	mPEG-SPA Tresylated mPEG	Shearwater Inc. Delgado et al, critical reviews in Therapeutic Drug Carrier Systems 9(3,4):249-304 (1992) WO 97/31022

		substituent		
-COOH	C-term, Asp, Glu	Polymer, e.g. PEG, with ester or amide group Oligosaccharide moiety	mPEG-Hz *In vitro* coupling	Shearwater Inc
-SH	Cys	Polymer, e.g. PEG, with disulfide, maleimide or vinyl sulfone group Oligosaccharide moiety	PEG-vinylsulphone PEG-maleimide *In vitro* coupling	Shearwater Inc Delgado et al, critical reviews in Therapeutic Drug Carrier Systems 9(3,4):249-304 (1992)
-OH	Ser, Thr, OH-, Lys	Oligosaccharide moiety PEG with ester, ether, carbamate, carbonate	*In vivo* O-linked glycosylation	
$-CONH_2$	Asn as part of an N-glycosylation site	Oligosaccharide moiety Polymer, e.g. PEG	*In vivo* N-glycosylation	
Aromatic residue	Phe, Tyr, Trp	Oligosaccharide moiety	*In vitro* coupling	
$-CONH_2$	Gln	Oligosaccharide moiety	*In vitro* coupling	Yan and Wold, Biochemistry, 1984, Jul 31; 23(16): 3759-65
Aldehyde Ketone	Oxidized oligosaccharide	Polymer, e.g. PEG,	PEG-hydrazide	Andresz et al., 1978, Makromol. Chem. 179:301, WO 92/16555, WO 00/23114
Guanidino	Arg	Oligosaccharide moiety	*In vitro* coupling	Lundblad and Noyes, Chimical Reagents for Protein Modification, CRC Press Inc. Boca Raton, FI
Imidazole ring	His	Oligosaccharide moiety	*In vitro* coupling	As for guanidine

For *in vivo* N-glycosylation, the term "attachment group" is used in an unconventional way to indicate the amino acid residues constituting an N-glycosylation site. Although the asparagine residue of the N-glycosylation site is where the oligosaccharide moiety is attached during glycosylation, such attachment cannot be achieved unless the other amino acid residues of the N-glycosylation site are present. Accordingly, when the non-peptide moiety is an oligosaccharide

moiety and the conjugation is to be achieved by N-glycosylation, the term "amino acid residue comprising an attachment group for the non-peptide moiety" as used in connection with alterations of the amino acid sequence of the polypeptide of interest is to be understood as meaning that one or more amino acid residues constituting an N-glycosylation site are to be altered in such a manner that either a functional N-glycosylation site is introduced into the amino acid sequence or removed from said sequence.

The term "comprising an attachment group" is intended to mean that the attachment group is present on an amino acid residue (including an N-glycosylation site) of the relevant peptide or polypeptide or on an oligosaccharide moiety attached to said peptide or polypeptide.

The term "contributing to an attachment group" as used in connection with the peptide addition X is intended to cover the situation, where an attachment group is formed from more than one amino acid residue (as is the case with an N-glycosylation site), and where at least one such amino acid residue originates from the peptide X and at least one amino acid residue originates from the polypeptide P, whereby the attachment group can be considered to bridge X and P (or, where relevant, P_x or P_y).

The term "non-structural part" as used about a part of the polypeptide P of interest is intended to indicate a part of either the C- or N-terminal end of the folded polypeptide (e.g. protein) that is outside the first structural element, such as an α-helix or a β-sheet structure. The non-structural part can easily be identified in a three-dimensional structure or model of the polypeptide. If no structure or model is available, a non-structural part typically comprises or consists of the first or last 1-20 amino acid residues, such as 1-10 amino acid residues of the amino acid sequence constituting the mature form of the polypeptide of interest.

Amino acid names and atom names (e.g. CA, CB, NZ, N, O, C, etc) are used as defined by the Protein DataBank (PDB) (www.pdb.org) which are based on the IUPAC nomenclature (IUPAC Nomenclature and Symbolism for Amino Acids and Peptides (residue names, atom names e.t.c.), *Eur. J. Biochem.*, **138**, 9-37 (1984) together with their corrections in *Eur. J. Biochem.*, **152**, 1 (1985). The term "amino acid residue" is intended to indicate an amino acid residue contained in the group consisting of alanine (Ala or A), cysteine (Cys or C), aspartic acid (Asp or D), glutamic acid (Glu or E), phenylalanine (Phe or F), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), lysine (Lys or K), leucine (Leu or L), methionine (Met or M), asparagine (Asn or N), proline (Pro or P), glutamine (Gln or Q), arginine (Arg or R), serine (Ser or S), threonine (Thr or T), valine (Val or V), tryptophan (Trp or W), and tyrosine (Tyr or Y) residues. The terminology used for identifying amino acid positions/mutations is illustrated as follows: A15 (indicates an alanine residue in position 15 of the polypeptide), A15T (indicates replacement of the alanine residue in position 15 with a threonine residue), A15T/S (indicates replacement of the alanine residue in position 15 with a threonine residue or a serine residue). Multiple substitutions are indicated with a "+", e.g. A15T+F57S means an amino acid sequence which comprises a substitution of the alanine residue in position 15 for a threonine residue and a substitution of the phenylalanine residue in position 57 for a serine residue.

The term "nucleotide sequence" is intended to indicate a consecutive stretch of two or more nucleotides. The nucleotide sequence can be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

"Cell", "host cell", "cell line" and "cell culture" are used interchangeably herein and all such terms should be understood to include progeny resulting from growth or culturing of a cell. "Transformation" and "transfection" are used interchangeably to refer to the process of introducing DNA into a cell.

"Operably linked" refers to the covalent joining of two or more nucleotide sequences in such a manner that the normal function of the sequences can be performed. For example, the nucleotide

sequence encoding a presequence or secretory leader is operably linked to a nucleotide sequence for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide: a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence.

"Introduction" or "removal" of an attachment group for the non-peptide moiety is normally achieved by introducing or removing an amino acid residue comprising such group to/from the relevant amino acid sequence, conveniently by suitable modification of the encoding nucleotide sequence. For instance, when an attachment group for a PEG molecule is to be introduced/removed, it will be understood that this be done by introducing/removing a codon for an amino acid residue, e.g. a lysine residue, comprising such group to/from the encoding nucleotide sequence. The term "introduce" is primarily intended to include substitution of an existing amino acid residue, but can also mean insertion of additional amino acid residue. The term "remove" is primarily intended to include substitution of the amino acid residue to be removed for another amino acid residue, but can also mean deletion (without substitution) of the amino acid residue to be removed.

The term "epitope" is used in its conventional meaning to indicate one or more amino acid residue(s) displaying specific 3D and/or charge characteristics at the surface of the polypeptide, which is/are capable of giving rise to an immune response in a mammal and/or specifically binding to an antibody raised against said epitope or which is/are capable of giving rise to an allergic response.

The term "unshielded epitope" is intended to indicate that the epitope is not shielded and therefore has the above properties. The term "shielded epitope" is intended to indicate that the non-peptide moiety shields, and thus inactivates the epitope, whereby it is no longer capable of giving rise to any substantial immune response in a mammal, e.g. due to inappropriate processing and/or presentation in the antigen presenting cells, and/or of reacting with an antibody raised against the unshielded epitope. The shielding should thus be effective in both the naïve mammal and mammals that already produce antibodies reacting with the unshielded epitope.

The degree of shielding of epitopes can be determined as reduced immunogenicity and/or reduced antibody reactivity and/or reduced reactivity with monoclonal antibodies raised against the epitope(s) in question using methods known in the art. The degree of shielding of allergenic epitopes can be determined, e.g., as described in WO 00/26354.

The term "reduced" as used about an immunogenic or allergic response is intended to indicate that a given molecule gives rise to a measurably lower immune or allergic response than a reference molecule, when determined under comparable conditions. Preferably, the relevant response is reduced by at least 25%, such as at least 50%, such as preferably by at least 75%, such as by at least 90% or even at least 100%.

The term "serum half-life" is used in its normal meaning, i.e. the time in which half of the relevant molecules circulate in the plasma or bloodstream prior to being cleared. Alternatively used terms include "plasma half-life", "circulating half-life", "serum clearance", "plasma clearance" and "clearance half-life". The term "functional *in vivo* half-life" is the time in which 50% of a given function (such as biological activity) of the relevant molecule is retained, when tested *in vivo* (such as the time at which 50% of the biological activity of the molecule is still present in the body/target organ, or the time at which the activity of the conjugate is 50% of the initial value). The molecule is normally cleared by the action of one or more of the reticuloendothelial systems (RES), kidney (e.g. by glomerular filtration), spleen or liver, or receptor-mediated elimination, or degraded by specific or unspecific proteolysis. Normally, clearance depends on size or hydrodynamic volume (relative to the cut-off for glomerular filtration), shape/rigidity, charge, attached carbohydrate chains, and the presence of cellular receptors for the molecule. The term "increased" as used about serum half-life or functional *in vivo* half-life is used to indicate that the relevant half-life of the relevant molecule is

statistically significantly increased relative to that of the reference molecule as determined under comparable conditions. For instance, the relevant half-life is increased by at least 25%, such as by at least 50%, by at least 100% or by at least 1000%.

The term "function" is intended to indicate one or more specific functions of the polypeptide of interest and is to be understood qualitatively (i.e. having a similar function as the polypeptide of interest) and not necessarily quantitatively (i.e. the magnitude of the function is not necessarily similar). Typically, a given polypeptide has many different functions, examples of which are given further below in the section entitled "Screening for or measurement of function". For therapeutically useful polypeptides an important "function" is biological activity, e.g. *in vitro* or *in vivo* bioactivity. For enzymes, an important function is biological activity such as catalytic activity.

The interchangeably used terms "measurable function" and "functional" are intended to indicate that the relevant function (preferably reflecting the intended use) of a conjugate of the invention is above detection limit when measured by standard methods known in the art, e.g. as an *in vitro* bioactivity and/or *in vivo* bioactivity. For instance, if the polypeptide is a hormone and the function of interest is the hormone's affinity towards a specific receptor a measurable function is defined to be a detectable affinity between the hormone conjugate and the receptor as determined by the normal methods used for measuring such affinity. If the polypeptide is an enzyme and a function of interest is the catalytic activity a measurable function is the enzyme conjugate's ability to catalyze a reaction involving the normal substrates for the enzyme as measured by the normal methods for determining the enzyme activity in question. Typically, if not otherwise stated herein, a measurable function is at least 2%, such as at least 5% of that of the un-conjugated polypeptide, as determined under comparable conditions, e.g. in the range of 2-1000%, such as 2-500% or 2-100%, such as 5-100% of that of the un-conjugated polypeptide.

The term "functional site" is intended to indicate one or more amino acid residues which is/are essential for or otherwise involved in the function or performance of the polypeptide, i.e. the amino acid residue(s) that mediate(s) a desired biological activity of the polypeptide in question. Such amino acid residues are "located at" the functional site. For instance, the functional site can be a binding site (e.g. a receptor-binding site of a hormone or growth factor or a ligand-binding site of a receptor), a catalytic site (e.g. of an enzyme), an antigen-binding site (e.g. of an antibody), a regulatory site (e.g. of a polypeptide subject to regulation), or an interaction site (e.g. for a regulatory protein or an inhibitor). The functional site can be determined by methods known in the art and is conveniently identified by analysing a three-dimensional or model structure of the polypeptide complexed to a relevant ligand.

The term "polypeptide" is intended to indicate any structural form (e.g. the primary, secondary or tertiary form (i.e. protein form)) of an amino acid sequence comprising more than 5 amino acid residues, which may or may not be post-translationally modified (e.g. acetylated, carboxylated, phosphorylated, lipidated, acylated or glycosylated (also in cases where the non-peptide moiety of a conjugate of the invention is different from an oligosaccharide moiety). The interchangeably used terms "native" and "wild-type" are used about a polypeptide which has an amino acid sequence that is identical to one found in nature. The native polypeptide is typically isolated from a naturally occurring source, in particular a mammalian or microbial source, such as a human source, or is produced recombinantly by use of a nucleotide sequence encoding the naturally occurring amino acid sequence. The term "native" is intended to encompass allelic variants of the polypeptide in question. A "variant" is a polypeptide, which has an amino acid sequence that differs from that of a native polypeptide in one or more amino acid residues. The variant is typically prepared by modification of a nucleotide sequence encoding the native polypeptide (e.g. to result in substitution, deletion or truncation of one or more amino acid residues of the polypeptide or by introduction (by addition or insertion) of one or more amino acid residues into the polypeptide) so

as to modify the amino acid sequence constituting said native polypeptide. A "fragment" is a part of a parent native or variant polypeptide, typically differing from such parent in one or more C-terminal or N-terminal amino acid residues or both types of such residues. Normally, the variant or fragment has retained at least one of the functions of the corresponding parent polypeptide (e.g. a biological function such as enzyme activity or receptor binding capability).

The term "antibody" includes single monoclonal antibodies (including agonist and antagonist antibodies) and antibody compositions with polyepitopic specificity (also termed polyclonal antibodies).

The term "monoclonal antibody" is used in its conventional meaning to indicate a population of substantially homogeneous antibodies. The individual antibodies comprised in the population have identical binding affinities and vary structurally only to a limited extent. Monoclonal antibodies are highly specific, being directed against a single epitope. Furthermore, in contrast to conventional (polyclonal) antibody preparations that typically include different antibodies directed against different epitopes, each monoclonal antibody is directed against a single epitope on the antigen. The antibody to be modified is preferably a human or humanized monoclonal antibody.

"Antibody fragment" is defined as a portion of an intact antibody comprising the antigen binding site or the entire or part of the variable region of the intact antibody, wherein the portion is free of the constant heavy chain domains (i.e. CH2, CH3, and CH4, depending on antibody isotype) of the Fc regions of the intact antibody. Examples of antibody fragments include Fab, Fab', Fab'-SH, F(ab')2, and Fv fragments; diabodies; any antibody fragment that is a polypeptide having a primary structure consisting of one uninterrupted sequence of contiguous amino acid residues (which may also be termed a single chain antibody fragment or a single chain polypeptide).

Conjugate of the invention

In its first aspect the invention relates to a conjugate of a polypeptide and a non-peptide moiety, wherein the polypeptide part of the conjugate comprises the primary structure,
NH_2- X-P-COOH or NH_2-P-X-COOH,
wherein
X is a peptide addition comprising or contributing to an attachment group for the non-peptide moiety, and P is a polypeptide of interest.

In one embodiment the polypeptide part of the conjugate consists essentially of or consists of a polypeptide with the primary structure NH_2- X-P-COOH or NH_2-P-X-COOH.

Usually, the peptide addition is fused to the N-terminal or C-terminal end of the polypeptide P as reflected in the above shown structure so as to provide an N- or C-terminal elongation of the polypeptide P. However, it is also possible to insert the peptide addition within the amino acid sequence of the polypeptide P. This is reflected in the conjugate according to the second aspect of the invention, which relates to a conjugate of a polypeptide and a non-peptide moiety, wherein the polypeptide part of the conjugate comprises the primary structure NH_2-P_x-X-P_y-COOH, wherein
P_x is an N-terminal part of a polypeptide P of interest,
P_y is a C-terminal part of said polypeptide P, and
X is a peptide addition comprising or contributing to an attachment group for the non-peptide moiety.

In one embodiment the polypeptide part of the conjugate consists essentially of or consists of a polypeptide with the primary structure NH_2-P_x-X-P_y-COOH.

In order to minimize structural changes effected by the insertion of the peptide addition within the sequence of the polypeptide P, it is desirable that it be inserted in a non-structural part

thereof. For instance, P_x is a non-structural N-terminal part of a mature polypeptide P, and P_y is a structural C-terminal part of said mature polypeptide, or P_x is a structural N-terminal part of a mature polypeptide P, and P_y is a non-structural C-terminal part of said mature polypeptide.

When the peptide addition comprises only few amino acid residues, e.g. 1-5 such as 1-3 amino acid residues, and in particular 1 amino acid residue, the peptide addition can be inserted into a loop structure of the polypeptide P and thereby elongate said loop.

Preferably, the conjugate of the invention has properties such as size, charge, molecular weight and/or hydrodynamic volume that are sufficient to reduce or escape clearance by any of the clearance mechanisms disclosed herein, in particular renal clerance.

In one embodiment, the conjugate of the invention has a molecular weight of at least 67 kDa, in particular at least 70 kDa as measured by SDS-PAGE according to Laemmli, U.K., Nature Vol 227 (1970), p680-85. This is of particular relevance when the polypeptide of interest is a therapeutically useful protein, the functional *in vivo* half-life of which is to be prolonged. A molecular weight of at least 67 kDa is obtainable by careful selection of the non-peptide moiety combined with the degree of conjugation to such moiety. For instance, for a polypeptide of interest having a molecular weight of at least 25 kDa linked to a peptide addition of 2 kDa, the combined extended polypeptide having at least two PEG-attachment groups, conjugation to two or more PEG molecules each having a molecular weight of 20 kDa results in a total molecular weight of at least 67 kDa.

Preferably, the conjugate of the invention has at least one of the following properties relative to the polypeptide P, the properties being measured under comparable conditions:
in vitro bioactivity which is at least 25% of that of the polypeptide P as measured under comparable conditions, increased affinity for a mannose receptor or other carbohydrate receptors, increased serum half-life, increased functional *in vivo* half-life, reduced renal clearance, reduced immunogenicity, increased resistance to proteolytic cleavage, improved targeting to lysosomes, macrophages and/or other subpopulations of human cells, improved stability in production, improved shelf life, improved formulation, e.g. liquid formulation, improved purification, improved solubility, and/or improved expression.
Improved properties are determined by conventional methods known in the art for determining such properties.

Polypeptide of interest
The present invention can be applied broadly. Thus, the polypeptide of interest can have any function and be of any origin. Accordingly, the polypeptide can be a protein, in particular a mature protein or a precursor form thereof or a functional fragment thereof that essentially has retained a biological activity of the mature protein. Furthermore, the polypeptide can be an oligopeptide that contains in the range of 30 to 4500 amino acids, preferably in the range of 40 to 3000 amino acids.

The polypeptide can be a native polypeptide or a variant thereof. For instance, the polypeptide is a variant that comprises at least one introduced and/or at least one removed attachment group for the non-peptide moiety as compared to the corresponding native polypeptide. The variant has retained at least one function of the corresponding native polypeptide, in particular a biological activity thereof.

The polypeptide can be a therapeutic polypeptide useful in human or veterinary therapy, i.e. a polypeptide that is physiologically active when introduced into the circulatory system of or otherwise administered to a human or an animal; a diagnostic polypeptide useful in diagnosis; or an industrial polypeptide useful for industrial purposes, such as in the manufacture of goods wherein the polypeptide constitutes a functional ingredient or wherein the polypeptide is used for processing or other modification of raw ingredients during the manufacturing process.

The polypeptide can be of mammalian origin, e.g. of human, porcine, ovine, urcine, murine, rabbit, donkey, or bat origin, of microbial origin, e.g. of fungal, yeast or bacterial origin, or can be derived from other sources such as venom, leech, frog or mosquito origin. Preferably, the industrial polypeptide of interest is of microbial origin and the therapeutic polypeptide of human origin.

Specific examples of groups of polypeptides to be modified according to the invention include: an antibody or antibody fragment, an immunoglobulin or immunoglobulin fragment, a plasma protein, an erythrocyte or thrombocyte protein, a cytokine, a growth factor, a profibrinolytic protein, a binding protein, a protease inhibitor, an antigen, an enzyme, a ligand, a receptor, or a hormone. Of particular interest is a polypeptide that mediates its biological effect by binding to a cellular receptor, when administered to a patient.

The antibody can be a polyclonal or monoclonal antibody, and can be of any origin including human, rabbit and murine origin. Preferably, the antibody is a human or humanized monoclonal antibody. Immunoglobulins of interest include IgG, IgE, IgM, IgA, and IgD and fragments thereof.

The non-antibody polypeptide of interest can be i) a plasma protein, e.g. a factor from the coagulation system, such as Factor VII, Factor VIII, Factor IX, Factor X, Factor XIII, thrombin, protein C, antithrombin III or heparin co-factor II, a factor from the fibrinolytic system such as pro-urokinase, urokinase, tissue plasminogen activator, plasminogen activator inhibitor 1 (PAI-1) or plasminogen activator inhibitor 2 (PAI-2), the Von Willebrand factor, or an α-1-proteinase inhibitor, ii) a erythrocyte or thrombocyte protein, e.g. hemoglobin, thrombospondin or platelet factor 4, iii) a cytokine, e.g. an interleukin such as IL-1 (e.g. IL-1α or IL-1β), IL-2, IL-4, IL-5, IL-6, IL-9, IL-10, IL-11, IL-12, IL-13, IL-15, IL-16, IL-17, IL-18, a cytokine-related polypeptide, such as IL-1Ra, an interferon such as interferon-α, interferon-β or interferon-γ, a colony-stimulating factor such as GM-CSF or G-CSF, stem cell factor (SCF), a binding protein, a member of the tumor necrosis factor family (e.g TNF-α, lymphotoxin-α, lymphotoxin-β, FasL, CD40L, CD30L, CD27L, Ox40L, 4-lBBL, RANKL, TRAIL, TWEAK, LIGHT, TRANCE, APRIL, THANK or TALL-1), iv) a growth factor, e.g platelet-derived growth factor (PDGF), transforming growth factor α (TGF-α), transforming growth factor β (TGF-β), epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), somatotropin (growth hormone), a somatomedin such as insulin-like growth factor I (IGF-I) or insulin-like growth factor II (IGF-II), erythropoietin (EPO), thrombopoietin (TPO) or angiopoietin, v) a profibrinolytic protein, e.g. staphylokinase or streptokinase, vi) a protease inhibitor, e.g. aprotinin or CI-2A, vii) an enzyme, e.g. superoxide dismutase, catalase, uricase, bilirubin oxidase, trypsin, papain, asparaginase, arginase, arginine deiminase, adenosin deaminase, ribonuclease, alkaline phosphatase, β-glucuronidase, purine nucleoside phosphorylase or batroxobin, viii) an opioid, e.g. endorphins, enkephalins or non-natural opioids, ix) a hormone or neuropeptide, e.g. insulin, calcitonin, glucagons, adrenocorticotropic hormone (ACTH), somatostatin, gastrins, cholecystokinins, parathyroid hormone (PTH), luteinizing hormone (LH), follicle-stimulating hormone (FSH), gonadotropin-releasing hormone, chorionic gonadotropin, corticotropin-releasing factor, vasopressin, oxytocin, antidiuretic hormones, thyroid-stimulating hormone, thyrotropin-releasing hormone, relaxin, glucagon-like peptide 1 (GLP-1), glucagon-like peptide 2 (GLP-2), prolactin, neuropeptide Y, peptide YY, pancreatic polypeptide, leptin, orexin, CART (cocaine and amphetamine regulated transcript), a CART-related peptide, melanocortins (melanocyte-stimulating hormones), melanin-concentrating hormone, natriuretic peptides, adrenomedullin, endothelin, exendin, secretin, amylin (IAPP;islet amyloid polypeptide precursor), vasoactive intestinal peptide (VIP), pituitary adenylate cyclase activating polypeptide (PACAP), agouti and agouti-related peptides or somatotropin-releasing hormones, or x) another type of protein or peptide such as thymosin, bombesin, bombesin-like peptides, heparin-binding protein, soluble

CD4, pigmentary hormones, hypothalamic releasing factor, malanotonins or phospholipase activating protein.

One group of polypeptides of particular interest in the present invention is selected from the group of lysosomal enzymes (as defined in US 5,929,304) such as those responsible for or otherwise involved in a lysosomal storage disease, i.e. enzymes that have a therapeutical effect on patients with a lysosomal storage disease. Such enzymes, e.g. include glucocerebrosidase, α-L-iduronidase, acid α-glucosidase, α-galactosidase, acid sphingomyelinase, and hexosaminidase. Also, other proteins involved in lysosomal storage diseases such as Saposin A, B, C or D (Nakano et al., J. Biochem. (Tokyo) 105, 152-154, 1989; Gavrieli-Rorman and Grabowski, Genomics 5, 486-492, 1989) can be modified as described herein. Preferably, these polypeptides are of human origin.

The present inventors have shown that providing such enzymes with additional N-linked sugar moieties considerably improve properties thereof, such as stability, targeting, expression, and *in vivo* activity and targeting. Accordingly, when the polypeptide P is such an enzyme, the conjugate is a glycosylated enzyme comprising one or more oligosaccharide moieties (constituting the non-peptide moiety) and one or more N-glycosylation site (constituting the attachment group).

The industrial polypeptide is typically an enzyme, in particular a microbial enzyme, and can be used in products or in the manufacture of products such as detergents, household articles, personal care products, agrochemicals, textile, food products, in particular bakery products, feed products, or in industrial processes such as hard surface cleaning. The industrial polypeptide is normally not intended for internal administration to humans or animals. Specific examples include hydrolases, such as proteases, lipases or cutinases, oxidoreductases, such as laccase and peroxidase, transferases such as transglutaminases, isomerases, such as protein disulphide isomerase and glucose isomerase, cell wall degrading enzymes such as cellulases, xylanases, pectinases, mannanases, etc., amylolytic enzymes such as endoamylases, e.g. alpha-amylases, or exo-amylases, e.g. beta-amylases or amyloglucosidases, etc. Further specific examples are those listed in WO 00/26354, the contents of which are incorporated herein by reference. Normally, an enzyme modified according to the present invention has one or more improved properties selected from the group consisting of increased stability (in particular against proteolytic degradation or thermal degradation) leading to, e.g., improved shelf life and improved performance in use; improved production, e.g. in terms of improved expression (e.g. as a consequence of improved secretion and/or increased stability of the expressed enzyme) and improved purification, decreased allergenicity, increased activity in the relevant industrial process in which it is used, and improved properties with respect to immobilization.

In one embodiment the conjugate comprises a personal care enzyme (i.e. an enzyme useful for personal care applications), which conjugate is incapable of passing the mucous membrane of a mammal in particular a human exposed to the conjugate. Thereby, allergenicity can be reduced or avoided. Furthermore, stability of such enzyme can be increased.

In another embodiment the conjugate comprises a lipase as disclosed in WO 97/04079, in particular a *Humicola lanuginosa* lipase, wherein the N- and/or C-terminal peptide addition, preferably N-terminal peptide addition comprises at least one attachment group for the non-peptide part of the conjugate. Thereby, the N- and/or C-terminal peptide addition is shielded from degradation and/or increased expression, including secretion, of the enzyme is likely to be obtained. In connection with this embodiment the N-terminal peptide addition can comprise any of the peptide additions disclosed in WO 97/04079.

In yet another embodiment the polypeptide P is an amyloglucosidase. The modification of such enzyme is contemplated to result in reduced or no degradation of the N-terminus of said enzyme (an otherwise well known problem associated with the recombinant production of

amyloglucosidase). In other words the N-terminus of the enzyme is protected by the non-peptide moiety attached to the N-terminal peptide addition of the amyloglucosidase.

In yet another embodiment the non-peptide moiety part of the conjugate is capable of cross-linking and thereby of being immobilized on a suitable solid support. Such cross-linking polymers are available from Shearwater Polymers, Inc. It will be understood that the peptide addition of the conjugate according to this embodiment comprises an attachment group for the cross-linking polymer in question. In connection with this embodiment the polypeptide P is preferably an enzyme, such as an industrial enzyme, e.g. an amyloglucosidase, an alpha-amylase, a glucose isomerase, an amidase, or a lipolytic enzyme.

Peptide addition

In principle the peptide addition X can be any stretch of amino acid residues ranging from a single amino acid residue to a mature protein. Usually, the peptide addition X comprises 1-500 amino acid residues, such as 2-500, normally 2-50 or 3-50 amino acid residues, such as 3-20 amino acid residues. The length of the peptide addition to be used for modification of a given polypeptide is dependent of or determined on the basis of a number of factors including the type of polypeptide of interest and the desired effect to be achieved by the modification. The peptide addition may be designed by a site-specific or random approach, e.g as out-lined in further detail in the Methods section below. This section also comprises a set of guidelines useful for preparing a peptide addition for use in the present invention are described. It will be understood that those guidelines are intended for illustration purposes only and that a person skilled in the art will be aware of alternative useful routes for design of peptide addition. Thus, the method of designing a peptide addition for use herein should not be considered limited to that described in the Materials section.

The number of attachment groups should be sufficient to provide the desired effect. Typically, the peptide addition X comprises 1-20, such as 1-10 attachment groups for the non-peptide moiety. For instance, the peptide addition X comprises 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 attachment groups for the non-peptide moiety. It is well known that one frequently occurring consequence of modifying an amino acid sequence of, e.g., a human protein is that new epitopes are created by such modification. In order to shield any new epitopes created by the peptide addition, it is desirable that sufficient attachment groups are present to enable shielding of all epitopes introduced into the sequence. This is e.g. achieved when the peptide addition X comprises at least one attachment group within a stretch of 30 contiguous amino acid residues, such as at least one attachment group within 20 amino acid residues or at least one attachment group within 10 amino acid residues, in particular 1-3 attachment groups within a stretch of 10 contiguous amino acid residues in the peptide addition X.

Thus, in one embodiment the peptide addition X comprises at least two attachment groups for the non-peptide moiety, wherein two of said amino acid residues are separated by at most 10 amino acid residues, none of which comprises an attachment group for the non-peptide moiety.

Furthermore, the polypeptide P part of a conjugate of the invention can comprise at least one introduced attachment group for the non-peptide moiety, in particular 1-5 introduced attachment groups. Analogously, the polypeptide P can comprise at least one removed attachment group for the non-peptide moiety, in particular 1-5 removed attachment groups.

Conjugate wherein non-peptide moiety is an oligosaccharide moiety

In one embodiment the non-peptide moiety is an oligosaccharide moiety and the attachment group is an *in vivo* glycosylation site, preferably an N-glycosylation site. Accordingly, the peptide addition X comprises at least one N-glycosylation site, typically at least two N-glycosylation sites. For instance, the peptide addition X has the structure X_1-N-X_2-T/S/C-Z, wherein X_1 is a peptide

comprising at least one amino acid residue or is absent, X_2 is any amino acid residue different from P, and Z is absent or a peptide comprising at least one amino acid residue. For instance, X_1 is absent, X_2 is an amino acid residue selected from the group consisting of I, A, G, V and S (all relatively small amino acid residues), and Z comprises at least 1 amino acid residue.

For instance, Z can be a peptide comprising 1-50 amino acid residues and, e.g., 1-10 glycosylation sites.

In another conjugate of the invention X_1 comprises at least one amino acid residue, e.g. 1-50 amino acid residues, X_2 is an amino acid residue selected from the group consisting of I, A, G, V and S, and Z is absent. For instance, X_1 comprises 1-10 glycosylation sites.

For instance, the peptide addition for use in the present invention can comprise a peptide sequence selected from the group consisting of INAT/S, GNIT/S, VNIT/S, SNIT/S, ASNIT/S, NIT/S, SPINAT/S, ASPINAT/S, ANIT/SANIT/SANI, and ANIT/SGSNIT/SGSNIT/S, wherein T/S is either a T or an S residue, preferably a T residue. The peptide addition can comprise one or more of these peptide sequences, i.e. at least two of said sequences either directly linked together or separated by one or more amino acid residues, or can contain two or more copies of any of these peptide sequence. It will be understood that the above specific sequences are given for illustrative purposes and thus do not constitute an exclusive list of peptide sequences of use in the present invention.

In a more specific embodiment the peptide addition X is selected from the group consisting of INAT/S, GNIT/S, VNIT/S, SNIT/S, ASNIT/S, NIT/S, SPINAT/S, ASPINAT/S, ANIT/SANIT/SANI, and ANIT/SGSNIT/SGSNIT/S, wherein T/S is either a T or an S residue, preferably a T residue.

As stated further above the polypeptide P forming part of a conjugate of the invention can be a native polypeptide that may or may not comprise one or more glycosylation sites. In order to further modify the glycosylation of the polypeptide P of interest (in terms of the number of oligosaccharide moieties attached to the polypeptide), the polypeptide P can be a variant of a native polypeptide that differs from said polypeptide in at least one introduced or at least one removed glycosylation site.

For instance, the polypeptide P comprises at least one introduced glycosylation site, in particular 1-5 introduced glycosylation sites, such as 2-5 introduced glycosylation sites.

In order to affect the total glycosylation of the polypeptide of interest the glycosylation site is introduced so that the N residue of said glycosylation site is exposed at the surface of the polypeptide, when folded in its active form. Likewise, a glycosylation site to be removed is selected from those having an N residue exposed at the surface of the polypeptide.

In one embodiment, the peptide addition X has an N residue in position –2 or –1, and the polypeptide P or P_x has a T or an S residue in position +1 or +2, respectively, the residue numbering being made relative to the N-terminal amino acid residue of P or P_x, whereby an N-glycosylation site is formed.

It will be understood that the conjugates described in the present section are glycosylated polypeptides. The section entitled "Conjugation to an oligosaccharide moiety" describes methods of preparing glycosylated polypeptides of the invention.

It can be advantageous that the conjugate (i.e.glycosylated polypeptide) further comprises at least one polymer molecule. For this purpose the polypeptide part of the conjugate must comprise at least one attachment group for the polymer molecule of choice. The attachment group can be located in the peptide addition X or the polypeptide P and in a position thereof that is exposed at the surface of the folded protein and thus accessible for conjugation to the polymer molecule. For instance, attachment to one or more polymer molecules increases the molecular weight of the polypeptide and can further serve to shield one or more epitopes thereof. The

polymer molecule may be any of the molecules mentioned in the section entitled "Conjugation to a polymer molecule", but is preferably selected from the group consisting of linear or branched polyethylene glycol or polyalkylene oxide. Most preferably, the polymer molecule is mPEG-SPA, mPEG-SCM, mPEG-BTC from Shearwater Polymers, Inc, SC-PEG from Enzon, Inc., tresylated mPEG (US 5,880,255) or oxycarbonyl-oxy-N-dicarboxyimide PEG (US 5,122,614) (and the relevant attachment group is one present on a lysine or N-terminal residue).

Other conjugates of the invention

In another embodiment the conjugate of the invention comprises a non-peptide moiety selected from the group consisting of a polymer molecule, a lipophilic compound and an organic derivatizing agent.

For this purpose the attachment group for the non-peptide moiety can be one present on an amino acid residue selected from the group consisting of the N-terminal amino acid residue of the polypeptide part of the conjugate, the C-terminal residue of the polypeptide part of the conjugate, lysine, cysteine, arginine, glutamine, aspartic acid, glutamic acid, serine, tyrosine, histidine, phenylalanine and tryptophan. For instance, the attachment group for the non-peptide moiety is an ε-amino group.

Analogously to what has been described above in the section entitled "Conjugate wherein the non-peptide moiety is an oligosaccharide moiety" the peptide addition X can comprise at least two attachment groups for the non-peptide moiety.

Also, the polypeptide P can be a variant of a native polypeptide, which as compared to said native polypeptide, comprises at least one introduced and/or at least one removed attachment group for the non-peptide moiety. For instance, the polypeptide P comprises at least one introduced attachment group, in particular 1-5 introduced attachment groups, such as 2-5 introduced attachment groups.

In addition to comprising a non-peptide moiety selected from the group consisting of a polymer molecule, a lipophilic compound and an organic derivatizing agent, the polypeptide part of the conjugate can be glycosylated. This requires that at least one of or possibly both of the polypeptide P and the peptide addition X comprises a glycosylation site that is accessible for glycosylation to be achieved. For instance, in addition to comprising an attachment group for the relevant of the above listed non-peptide moieties, peptide addition X can comprise at least one glycosylation site. The glycosylation site can be an *in vivo* or an *in vitro* glycosylation site, preferably an N-glycosylation site.

Thus, the conjugate of the invention can comprise more than one type of non-peptide moiety, i.e. two or more types of non-peptide moieties, e.g. two types of non-peptide moieties. The peptide addition may comprise an attachment group for both or all such non-peptide moieties or only for one of the types (attachment group(s) for the other type(s) being provided by the polypeptide of interest). For instance, the conjugate of the invention can comprise one or more N-linked oligosaccharide moieties and one or more polymer molecules of the polyalkylene oxide type, or one or more lipophilic groups. The conjugation to two or more different non-peptide moieties can be done simultaneous or sequentially.

Methods of preparing a conjugate of the invention

In the following sections "Conjugation to a lipophilic compound", "Conjugation to a polymer molecule", "Conjugation to an oligosaccharide moiety" and "Conjugation to an organic derivatizing agent" conjugation to specific types of non-peptide moieties is described.

It will be understood that a conjugation step of any method of the invention only finds relevance when a non-polypeptide moiety other than an *in vivo* attached oligosaccharide moiety is

to be conjugated to the polypeptide, since *in vivo* glycosylation takes place during the expression step when using an appropriate glycosylating host cell as expression host. Accordingly, whenever a conjugation step occurs in the present invention this is intended to be conjugation to a non-polypeptide moiety other than an oligosaccharide moiety attached by *in vivo* glycosylation during expression in a glycosylating organism.

Conjugation to a lipophilic compound

The polypeptide part of the conjugate and the lipophilic compound can be conjugated to each other, either directly or by use of a linker. The lipophilic compound can be a natural compound such as a saturated or unsaturated fatty acid, a fatty acid diketone, a terpene, a prostaglandin, a vitamine, a carotenoide or steroide, or a synthetic compound such as a carbon acid, an alcohol, an amine and sulphonic acid with one or more alkyl-, aryl-, alkenyl- or other multiple unsaturated compounds. Furthermore, the lipophilic compound may be any of the lipophilic substituents disclosed in WO 97/31022, the contents of which are incorporated herein by reference. The conjugation between the polypeptide and the lipophilic compound, optionally through a linker can be done according to methods known in the art, e.g. as described by Bodanszky in Peptide Synthesis, John Wiley, New York, 1976 and in WO 96/12505 and further as described in WO 97/31022.

Conjugation to a polymer molecule

The polymer molecule to be coupled to the polypeptide part of a conjugate of the invention can be any suitable polymer molecule, such as a natural or synthetic homo-polymer or heteropolymer, typically with a molecular weight in the range of 300-100,000 Da, such as 300-20,000 Da, more preferably in the range of 500-10,000 Da, even more preferably in the range of 500-5000 Da.

Examples of homo-polymers include a polyol (i.e. poly-OH), a polyamine (i.e. poly-NH_2) and a polycarboxylic acid (i.e. poly-COOH). A hetero-polymer is a polymer that comprises different coupling groups, such as a hydroxyl group and an amine group.

Examples of suitable polymer molecules include polymer molecules selected from the group consisting of polyalkylene oxide (PAO), including polyalkylene glycol (PAG), such as polyethylene glycol (PEG) and polypropylene glycol (PPG), branched PEGs, poly-vinyl alcohol (PVA), poly-carboxylate, poly-(vinylpyrolidone), polyethylene-co-maleic acid anhydride, polystyrene-co-malic acid anhydride, dextran, including carboxymethyl-dextran, or any other biopolymer suitable for the intended purpose, such as for reducing immunogenicity and/or increasing functional *in vivo* half-life and/or serum half-life, or for providing immobilization properties to the polypeptide (as discussed in the section entitled "Polypeptide of interest". Another example of a polymer molecule is human albumin or another abundant plasma protein. Generally, polyalkylene glycol-derived polymers are biocompatible, non-toxic, non-antigenic, non-immunogenic, have various water solubility properties, and are easily excreted from living organisms.

PEG is the preferred polymer molecule for reducing immunogenicity, allergenicity and/or increasing half-life, since it has only few reactive groups capable of cross-linking compared, e.g., to polysaccharides such as dextran, and the like. In particular, monofunctional PEG, e.g. methoxypolyethylene glycol (mPEG), is of interest since its coupling chemistry is relatively simple (only one reactive group is available for conjugating with attachment groups on the polypeptide). Consequently, the risk of cross-linking is eliminated, the resulting polypeptide conjugates are more homogeneous and the reaction of the polymer molecules with the polypeptide is easier to control.

To effect covalent attachment of the polymer molecule(s) to the polypeptide part of the conjugate, the hydroxyl end groups of the polymer molecule must be provided in activated form, i.e. with reactive functional groups. Suitable activated polymer molecules are commercially available, e.g. from Shearwater Polymers, Inc., Huntsville, AL, USA. Alternatively, the polymer

molecules can be activated by conventional methods known in the art, e.g. as disclosed in WO 90/13540. Specific examples of activated linear or branched polymer molecules for use in the present invention are described in the Shearwater Polymers, Inc. 1997 and 2000 Catalogs (Functionalized Biocompatible Polymers for Research and pharmaceuticals, Polyethylene Glycol and Derivatives, incorporated herein by reference). Specific examples of activated PEG polymers include the following linear PEGs: NHS-PEG (e.g. SPA-PEG, SSPA-PEG, SBA-PEG, SS-PEG, SSA-PEG, SC-PEG, SG-PEG, and SCM-PEG), and NOR-PEG), BTC-PEG, EPOX-PEG, NCO-PEG, NPC-PEG, CDI-PEG, ALD-PEG, TRES-PEG, VS-PEG, IODO-PEG, and MAL-PEG, and branched PEGs such as PEG2-NHS and those disclosed in US 5,932,462 and US 5,643,575, both of which are incorporated herein by reference. Furthermore, the following publications, incorporated herein by reference, disclose useful polymer molecules and/or PEGylation chemistries: US 5,824,778, US 5,476,653, WO 97/32607, EP 229,108, EP 402,378, US 4,902,502, US 5,281,698, US 5,122,614, US 5,219,564, WO 92/16555, WO 94/04193, WO 94/14758, WO 94/17039, WO 94/18247, WO 94/28024, WO 95/00162, WO 95/11924, WO95/13090, WO 95/33490, WO 96/00080, WO 97/18832, WO 98/41562, WO 98/48837, WO 99/32134, WO 99/32139, WO 99/32140, WO 96/40791, WO 98/32466, WO 95/06058, EP 439 508, WO 97/03106, WO 96/21469, WO 95/13312, EP 921 131, US 5,736,625, WO 98/05363, EP 809 996, US 5,629,384, WO 96/41813, WO 96/07670, US 5,473,034, US 5,516,673, EP 605 963, US 5,382,657, EP 510 356, EP 400 472, EP 183 503 and EP 154 316.

The conjugation of the polypeptide part of the conjugate and the activated polymer molecules is conducted by use of any conventional method, e.g. as described in the following references (which also describe suitable methods for activation of polymer molecules): R.F. Taylor, (1991), "Protein immobilisation. Fundamental and applications", Marcel Dekker, N.Y.; S.S. Wong, (1992), "Chemistry of Protein Conjugation and Crosslinking", CRC Press, Boca Raton; G.T. Hermanson et al., (1993), "Immobilized Affinity Ligand Techniques", Academic Press, N.Y.). The skilled person will be aware that the activation method and/or conjugation chemistry to be used depends on the attachment group(s) of the polypeptide (examples of which are given further above), as well as the functional groups of the polymer (e.g. being amine, hydroxyl, carboxyl, aldehyde, sulfydryl, succinimidyl, maleimide, vinysulfone or haloacetate). The PEGylation can be directed towards conjugation to all available attachment groups on the polypeptide (i.e. such attachment groups that are exposed at the surface of the polypeptide) or can be directed towards one or more specific attachment groups, e.g. the N-terminal amino group (US 5,985,265). Furthermore, the conjugation can be achieved in one step or in a stepwise manner (e.g. as described in WO 99/55377).

It will be understood that the PEGylation is designed so as to produce the optimal molecule with respect to the number of PEG molecules attached, the size and form of such molecules (e.g. whether they are linear or branched), and where in the polypeptide such molecules are attached. For instance, the molecular weight of the polymer to be used can be chosen on the basis of the desired effect to be achieved. For instance, if the primary purpose of the conjugation is to achieve a conjugate having a high molecular weight (e.g. to reduce renal clearance) it is usually desirable to conjugate as few high Mw polymer molecules as possible to obtain the desired molecular weight. When a high degree of epitope shielding is desirable this can be obtained by use of a sufficiently high number of low molecular weight polymer molecules (e.g. with a molecular weight of about 5,000 Da) to effectively shield all or most epitopes of the polypeptide. For instance, 2-8, such as 3-6 such polymers can be used.

In connection with conjugation to only a single attachment group on the protein (as described in US 5,985,265), it can be advantageous that the polymer molecule, which can be linear or branched, has a high molecular weight, e.g. about 20 kDa.

17

Normally, the polymer conjugation is performed under conditions aiming at reacting all available polymer attachment groups with polymer molecules. Typically, the molar ratio of activated polymer molecules to polypeptide is up to about 1000-1, in particular 200-1, preferably 100-1, such as 10-1 or 5-1, but also equimolar ratios can be used in order to obtain optimal reaction.

It is also contemplated according to the invention to couple the polymer molecules to the polypeptide through a linker. Suitable linkers are well known to the skilled person. A preferred example is cyanuric chloride (Abuchowski et al., (1977), J. Biol. Chem., 252, 3578-3581; US 4,179,337; Shafer et al., (1986), J. Polym. Sci. Polym. Chem. Ed., 24, 375-378.

Subsequent to the conjugation residual activated polymer molecules are blocked according to methods known in the art, e.g. by addition of primary amine to the reaction mixture, and the resulting inactivated polymer molecules are removed by a suitable method.

In a specific embodiment, the polypeptide conjugate of the invention is one that comprises one or more PEG molecules attached to the peptide addition, but not to the polypeptide P. For instance, the PEG molecule is attached to one or more cysteine residues present in the peptide addition X and, if necessary, one or more cysteine residues have been removed from the polypeptide P of interest in order to avoid conjugation thereto. The polypeptide according to this embodiment can further comprise one or more oligosaccharide moieties attached to a glycosylation site of the polypeptide (present in either or both of the polypeptide P and peptide addition X).

In another specific embodiment, the polypeptide conjugate of the invention comprises at least one PEG molecule attached to a lysine residue of the peptide addition X, in particular a linear or branched PEG molecule with a molecular weight of at least 5kDa.

Methods of preparing a polypeptide of the invention

The invention further comprises a method of producing the polypeptide part of a conjugate of the invention, including a glycosylated polypeptide of the invention, which method comprises culturing a host cell transformed or transfected with a nucleotide sequence encoding the polypeptide under conditions permitting the expression of the polypeptide, and recovering the polypeptide from the culture.

Apart from recombinant production, polypeptides of the invention may be produced, albeit less efficiently, by chemical synthesis or a combination of chemical synthesis and recombinant DNA technology.

The nucleotide sequence of the invention encoding a polypeptide part of a conjugate of the invention may be constructed by isolating or synthesizing a nucleotide sequence encoding the parent polypeptide and fusing a nucleotide sequence encoding the relevant peptide addition in accordance with established technologies. To the extent amino acid modifications are to be made in the parent polypeptide, these are conveniently done by mutagenesis, e.g. using site-directed mutagenesis in accordance with well-known methods, e.g. as described in Nelson and Long, Analytical Biochemistry 180, 147-151, 1989, random mutagenesis or shuffling.

The nucleotide sequence may be prepared by chemical synthesis, e.g. by using an oligonucleotide synthesizer, wherein oligonucleotides are designed based on the amino acid sequence of the desired polypeptide, and preferably selecting those codons that are favoured in the host cell in which the recombinant polypeptide will be produced. For example, several small oligonucleotides coding for portions of the desired polypeptide may be synthesized and assembled by polymerase chain reaction (PCR), ligation or ligation chain reaction (LCR). The individual oligonucleotides typically contain 5' or 3' overhangs for complementary assembly.

Once assembled (by synthesis, site-directed mutagenesis or another method), the nucleotide sequence encoding the polypeptide part of the conjugate may be inserted into a recombinant vector

and operably linked to control sequences necessary for expression of thereof in the desired transformed host cell.

It should of course be understood that not all vectors and expression control sequences function equally well to express the nucleotide sequence encoding the polypeptide part of a conjugate of the invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation. For example, in selecting a vector, the host must be considered because the vector must replicate in it or be able to integrate into the chromosome. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleotide sequence encoding the polypeptide, particularly as regards potential secondary structures. Hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleotide sequence, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification of the products coded for by the nucleotide sequence.

The recombinant vector may be an autonomously replicating vector, i.e. a vector existing as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector is one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector, in which the nucleotide sequence encoding the polypeptide part of a conjugate of the invention is operably linked to additional segments required for transcription of the nucleotide sequence. The vector is typically derived from plasmid or viral DNA. A number of suitable expression vectors for expression in the host cells mentioned herein are commercially available or described in the literature. Useful expression vectors for eukaryotic hosts, include, for example, vectors comprising expression control sequences from SV40, bovine papilloma virus, adenovirus and cytomegalovirus. Specific vectors are, e.g., pCDNA3.1(+)\Hyg (Invitrogen, Carlsbad, CA, USA) and pCI-neo (Stratagene, La Jolla, CA, USA). Useful expression vectors for yeast cells include the 2μ plasmid and derivatives thereof, the POT1 vector (US 4,931,373), the pJSO37 vector described in (Okkels, Ann. New York Acad. Sci. 782, 202-207, 1996) and pPICZ A, B or C (Invitrogen, Carlsbad, CA, USA). Useful vectors for insect cells include pVL941, pBG311 (Cate et al., "Isolation of the Bovine and Human Genes for Mullerian Inhibiting Substance And Expression of the Human Gene In Animal Cells", Cell, 45, pp. 685-98 (1986), pBluebac 4.5 and pMelbac (both available from Invitrogen, Carlsbad, CA, USA).

Other vectors for use in this invention include those that allow the nucleotide sequence encoding the polypeptide part of a conjugate of the invention to be amplified in copy number. Such amplifiable vectors are well known in the art. They include, for example, vectors able to be amplified by DHFR amplification (see, e.g., Kaufman, U.S. Pat. No. 4,470,461, Kaufman and Sharp, "Construction Of A Modular Dihydrafolate Reductase cDNA Gene: Analysis Of Signals Utilized For Efficient Expression", Mol. Cell. Biol., 2, pp. 1304-19 (1982)) and glutamine synthetase ("GS") amplification (see, e.g., US 5,122,464 and EP 338,841).

The recombinant vector may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. An example of such a sequence (when the host cell is a mammalian cell) is the SV40 origin of replication. When the host cell is a yeast cell, suitable sequences enabling the vector to replicate are the yeast plasmid 2μ replication genes REP 1-3 and origin of replication.

The vector may also comprise a selectable marker, e.g. a gene the product of which complements a defect in the host cell, such as the gene coding for dihydrofolate reductase (DHFR) or the Schizosaccharomyces pombe TPI gene (described by P.R. Russell, Gene 40, 1985, pp. 125-130), or one which confers resistance to a drug, e.g. ampicillin, kanamycin, tetracyclin, chloramphenicol, neomycin, hygromycin or methotrexate. For filamentous fungi, selectable markers include amdS, pyrG, arcB, niaD, sC.

The term "control sequences" is defined herein to include all components, which are necessary or advantageous for the expression of the polypeptide part of a conjugate of the invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, enhancer or upstream activating sequence, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter operably linked to the nucleotide sequence encoding the polypeptide.

"Operably linked" refers to the covalent joining of two or more nucleotide sequences, by means of enzymatic ligation or otherwise, in a configuration relative to one another such that the normal function of the sequences can be performed. For example, the nucleotide sequence encoding a presequence or secretory leader is operably linked to a nucleotide sequence for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide: a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the nucleotide sequences being linked are contiguous and, in the case of a secretory leader, contiguous and in reading phase. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, then synthetic oligonucleotide adaptors or linkers are used, in conjunction with standard recombinant DNA methods.

A wide variety of expression control sequences may be used in the present invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors as well as any sequence known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

Examples of suitable control sequences for directing transcription in mammalian cells include the early and late promoters of SV40 and adenovirus, e.g. the adenovirus 2 major late promoter, the MT-1 (metallothionein gene) promoter, the human cytomegalovirus immediate-early gene promoter (CMV), the human elongation factor 1α (EF-1α) promoter, the *Drosophila* minimal heat shock protein 70 promoter, the Rous Sarcoma Virus (RSV) promoter, the human ubiquitin C (UbC) promoter, the human growth hormone terminator, SV40 or adenovirus Elb region polyadenylation signals and the Kozak consensus sequence (Kozak, M. *J Mol Biol* 1987 Aug 20;196(4):947-50).

In order to improve expression in mammalian cells a synthetic intron may be inserted in the 5' untranslated region of the nucleotide sequence encoding the polypeptide of the invention. An example of a synthetic intron is the synthetic intron from the plasmid pCI-Neo (available from Promega Corporation, WI, USA).

Examples of suitable control sequences for directing transcription in insect cells include the polyhedrin promoter, the P10 promoter, the *Autographa californica* polyhedrosis virus basic protein promoter, the baculovirus immediate early gene 1 promoter and the baculovirus 39K delayed-early gene promoter, and the SV40 polyadenylation sequence.

Examples of suitable control sequences for use in yeast host cells include the promoters of the yeast α-mating system, the yeast triose phosphate isomerase (TPI) promoter, promoters from

yeast glycolytic genes or alcohol dehydogenase genes, the ADH2-4c promoter and the inducible GAL promoter.

Examples of suitable control sequences for use in filamentous fungal host cells include the ADH3 promoter and terminator, a promoter derived from the genes encoding *Aspergillus oryzae* TAKA amylase triose phosphate isomerase or alkaline protease, an *A. niger* α-amylase, *A. niger* or *A. nidulans* glucoamylase, *A. nidulans* acetamidase, *Rhizomucor miehei* aspartic proteinase or lipase, the TPI1 terminator and the ADH3 terminator.

The nucleotide sequence of the invention may or may not also include a nucleotide sequence that encode a signal peptide. The signal peptide is present when the polypeptide is to be secreted from the cells in which it is expressed. Such signal peptide, if present, should be one recognized by the cell chosen for expression of the polypeptide. The signal peptide may be homologous (e.g. be that normally associated with the parent polypeptide in question) or heterologous (i.e. originating from another source than the parent polypeptide) to the polypeptide or may be homologous or heterologous to the host cell, i.e. be a signal peptide normally expressed from the host cell or one which is not normally expressed from the host cell. Accordingly, the signal peptide may be prokaryotic, e.g. derived from a bacterium, or eukaryotic, e.g. derived from a mammalian, or insect, filamentous fungal or yeast cell.

The presence or absence of a signal peptide will, e.g., depend on the expression host cell used for the production of the polypeptide, the protein to be expressed (whether it is an intracellular or extracelluar protein) and whether it is desirable to obtain secretion. For use in filamentous fungi, the signal peptide may conveniently be derived from a gene encoding an *Aspergillus* sp. amylase or glucoamylase, a gene encoding a *Rhizomucor miehei* lipase or protease or a *Humicola lanuginosa* lipase. The signal peptide is preferably derived from a gene encoding *A. oryzae* TAKA amylase, *A. niger* neutral α-amylase, *A. niger* acid-stable amylase, or *A. niger* glucoamylase. For use in insect cells, the signal peptide may conveniently be derived from an insect gene (cf. WO 90/05783), such as the lepidopteran Manduca sexta adipokinetic hormone precursor, (cf. US 5,023,328), the honeybee melittin (Invitrogen, Carlsbad, CA, USA), ecdysteroid UDPglucosyltransferase (egt) (Murphy et al., Protein Expression and Purification 4, 349-357 (1993) or human pancreatic lipase (hpl) (Methods in Enzymology 284, pp. 262-272, 1997).

Specific examples of signal peptides for use in mammalian cells include that of human glucocerebrosidase apparent from the examples hereinafter or the murine Ig kappa light chain signal peptide (Coloma, M (1992) J. Imm. Methods 152:89-104). For use in yeast cells suitable signal peptides have been found to be the α-factor signal peptide from *S. cereviciae*. (cf. US 4,870,008), the signal peptide of mouse salivary amylase (cf. O. Hagenbuchle et al., Nature 289, 1981, pp. 643-646), a modified carboxypeptidase signal peptide (cf. L.A. Valls et al., Cell 48, 1987, pp. 887-897), the yeast BAR1 signal peptide (cf. WO 87/02670), and the yeast aspartic protease 3 (YAP3) signal peptide (cf. M. Egel-Mitani et al., Yeast 6, 1990, pp. 127-137).

Any suitable host may be used to produce the polypeptide part of a conjugate of the invention, including bacteria, fungi (including yeasts), plant, insect, mammal, or other appropriate animal cells or cell lines, as well as transgenic animals or plants. When a non-glycosylating organism such as *E. coli* is used, and the polypeptide is to be a glycosylated polypeptide, the expression in *E. coli* is preferably followed by suitable *in vitro* glycosylation.

Examples of bacterial host cells include grampositive bacteria such as strains of *Bacillus*, e.g. *B. brevis* or *B. subtilis*, *Pseudomonas* or *Streptomyces*, or gramnegative bacteria, such as strains of *E. coli*. The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast transformation (see, *e.g.*, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, *e.g.*, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221),

electroporation (see, *e.g.*, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, *e.g.*, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278).

Examples of suitable filamentous fungal host cells include strains of *Aspergillus*, e.g. *A. oryzae*, *A. niger*, or *A. nidulans*, *Fusarium* or *Trichoderma*. Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and US 5,679,543. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156 and WO 96/00787. Yeast may be transformed using the procedures described by Becker and Guarente, *In* Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology*, *Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; and Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920.

When the polypeptide part of a conjugate of the invention is to be *in vivo* glycosylated, the host cell is selected from a group of host cells capable of generating the desired glycosylation of the polypeptide. Thus, the host cell may advantageously be selected from a yeast cell, insect cell, or mammalian cell.

Examples of suitable yeast host cells include strains of *Saccharomyces*, e.g. *S. cerevisiae*, *Schizosaccharomyces*, *Klyveromyces*, *Pichia*, such as *P. pastoris* or *P. methanolica*, *Hansenula*, such as *H. polymorpha* or *yarrowia*. Of particular interest are yeast glycosylation mutant cells, e.g. derived from *S. cereviciae*, *P. pastoris* or *Hansenula* spp. (e.g. the *S. cereviciae* glycosylation mutants och1, ochi mnm1 or och1 mnm1 alg3 described by Nagasu et al. Yeast 8, 535-547, 1992 and Nakanisho-Shindo et al. J. Biol. Chem. 268, 26338-26345, 1993). Methods for transforming yeast cells with heterologous DNA and producing heterologous polypeptides therefrom are disclosed by Clontech Laboratories, Inc, Palo Alto, CA, USA (in the product protocol for the Yeastmaker™ Yeast Tranformation System Kit), and by Reeves et al., FEMS Microbiology Letters 99 (1992) 193-198, Manivasakam and Schiestl, Nucleic Acids Research, 1993, Vol. 21, No. 18, pp. 4414-4415 and Ganeva et al., FEMS Microbiology Letters 121 (1994) 159-164.

Examples of suitable insect host cells include a *Lepidoptora* cell line, such as *Spodoptera frugiperda* (Sf9 or Sf21) or *Trichoplusia ni* cells (High Five) (US 5,077,214). Transformation of insect cells and production of heterologous polypeptides therein may be performed as described by Invitrogen, Carlsbad, CA, USA.

Examples of suitable mammalian host cells include Chinese hamster ovary (CHO) cell lines, (e.g. CHO-K1; ATCC CCL-61), Green Monkey cell lines (COS) (e.g. COS 1 (ATCC CRL-1650), COS 7 (ATCC CRL-1651)); mouse cells (e.g. NS/O), Baby Hamster Kidney (BHK) cell lines (e.g. ATCC CRL-1632 or ATCC CCL-10), and human cells (e.g. HEK 293 (ATCC CRL-1573)), as well as plant cells in tissue culture. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Maryland. Of interest for the present purpose are a mammalian glycosylation mutant cell line, such as CHO-LEC1, CHOL-LEC2 or CHO-LEC18 (CHO-LEC1: Stanley et al. Proc. Natl. Acad. USA 72, 3323-3327, 1975 and Grossmann et al., J. Biol. Chem. 270, 29378-29385, 1995, CHO-LEC18: Raju et al. J. Biol. Chem. 270, 30294-30302, 1995).

Methods for introducing exogeneous DNA into mammalian host cells include calcium phosphate-mediated transfection, electroporation, DEAE-dextran mediated transfection, liposome-mediated transfection, viral vectors and the transfection method described by Life Technologies Ltd, Paisley, UK using Lipofectamin 2000. These methods are well known in the art and e.g. described by Ausbel et al. (eds.), 1996, Current Protocols in Molecular Biology, John Wiley & Sons, New York, USA. The cultivation of mammalian cells are conducted according to established methods, e.g. as disclosed in (Animal Cell Biotechnology, Methods and Protocols, Edited by Nigel

Jenkins, 1999, Human Press Inc, Totowa, New Jersey, USA and Harrison MA and Rae IF, General Techniques of Cell Culture, Cambridge University Press 1997).

In the production methods of the present invention, cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, cells are cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermenters performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The resulting polypeptide may be recovered by methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray drying, evaporation, or precipitation.

The polypeptides may be purified by a variety of procedures known in the art including, but not limited to, chromatography (*e.g.*, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (*e.g.*, preparative isoelectric focusing), differential solubility (*e.g.*, ammonium sulfate precipitation), SDS-PAGE, or extraction (see, *e.g.*, *Protein Purification*, J-C Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Other methods of the invention

In accordance with a specific aspect a nucleotide sequence encoding the polypeptide part of a conjugate of the invention is prepared by a method comprising
a) subjecting a nucleotide sequence encoding the polypeptide P to elongation mutagenesis,
b) expressing the mutated nucleotide sequence obtained in step a),
c) conjugating polypeptides expressed in step b) to the non-peptide moiety to be used for preparing the relevant polypeptide conjugate,
d) selecting conjugates comprising at least one non-peptide moiety attached to the peptide addition part of the polypeptide, and
e) isolating a nucleotide sequence encoding the polypeptide part of conjugates selected in step d).

In the present context the term "elongation mutagenesis" is intended to indicate any manner in which the nucleotide sequence encoding the parent polypeptide P can be extended to further encode the peptide addition For instance, a nucleotide sequence encoding a peptide addition of a suitable length may be synthesized and fused to a nucleotide sequence encoding the polypeptide P. The resulting fused nucleotide sequence may then be subjected to further modification by any suitable method, e.g. one which involves gene shuffling, other recombination between nucleotide sequences, random mutagenesis, random elongation mutagenesis or any combination of these methods. Such methods are further described in the Methods section herein.

The expression and conjugation steps are conducted as described in further detail elsewhere in the present application, and the selection step d) using any suitable method available in the art.

In one embodiment the above method further comprises screening conjugates resulting from step c) for at least one improved property, in particular any of those improved properties listed herein, one prior to the selection step, and wherein the selection step d) further comprises selecting conjugates having such improved property.

Furthermore, in the above method the elongation mutagenesis can be conducted so as to enrich for codons encoding an amino acid residue comprising an attachment group for the non-peptide moiety, in particular an *in vivo* glycosylation site.

Still further, the above method can comprise subjecting the part of the nucleotide sequence encoding the polypeptide P of interest to mutagenesis to remove and/or introduce amino acid residues comprising attachment groups for the non-peptide moiety. The nucleotide sequence may be subjected to any type of mutagenesis, e.g. any of those described herein. The mutagenesis of the nucleotide sequence encoding the polypeptide P of interest can be conducted prior to assembling the sequence with that encoding the peptide addition, concomitantly with or after any mutagenesis of the peptide addition part of the assembled nucleotide sequence.

In a further aspect, the invention relates to a method of producing a glycosylated polypeptide encoded by a nucleotide sequence of the invention prepared by the above method, wherein the nucleotide sequence encoding the polypeptide selected in step c) is expressed in a glycosylating host cell and the resulting glycosylated expressed polypeptide is recovered.

In a still further aspect the invention relates to a method of improving one or more selected properties of a polypeptide P of interest, which method comprises

a) preparing a nucleotide sequence encoding a polypeptide comprising or consisting essentially of the primary structure

NH_2-X-P-COOH or NH_2-P-X-COOH,

wherein
X is a peptide addition comprising or contributing to an attachment group for a non-peptide moiety that is capable of conferring the selected improved property/ies to the polypeptide P, when conjugated thereto,
b) expressing the nucleotide sequence of a) in an suitable host cell,
c) conjugating the expressed polypeptide of b) to the non-peptide moiety, and
d) recovering the conjugate resulting from step c).

For instance, the polypeptide is any of those constituting the polypeptide of a conjugate of the invention. For instanc,e the nucleotide sequence of step a) is prepared by subjecting a nucleotide sequence encoding the polypeptide P to elongation mutagenesis, e.g. to enrich for codons encoding an amino acid residue comprising an attachment group for the non-peptide moiety, in particular an *in vivo* glycosylation site. Also, in the preparation of the nucleotide sequence of a), the part of the nucleotide sequence encoding the polypeptide P can be subjected to mutagenesis to remove and/or introduce an attachment group for the non-peptide moiety.

The method according to this aspect can further comprise a screening step (after step c)), wherein the conjugate resulting from step c) is screened for one or more improved properties, in particular any of those improved properties which are described hereinabove.

When the non-peptide moiety is a sugar moiety, the host cell in step b) can be a glycosylating host cell, and the conjugation in step c) is achieved by *in vivo* glycosylation during the expression step b).

Usually, when a polypeptide conjugate has been selected in a screening step of a method of the invention the nucleotide sequence encoding the polypeptide part of the conjugate is isolated and used for expression of larger amounts of the polypeptide. The amino acid sequence of the resulting polypeptide is determined and the polypeptide is subjected to conjugation in a larger scale. Subsequently, the polypeptide conjugate is assayed with respect to the property to be improved.

Uses of a polypeptide conjugate of the invention

It will be understood that polypeptide conjugates of the invention can be used for a variety of purposes, depending on the type and nature of polypeptide. For instance, it is contemplated that a conjugate of the invention prepared from a therapeutic polypeptide is useful for the same therapeutic purposes as the parent polypeptide, i.e. for the treatment of a particular disease. Accordingly, the polypeptide conjugate of the invention may be formulated into a pharmaceutical composition. Also, when the conjugate of the invention is an *in vivo* glycosylated polypeptide which does not comprise any other type of non-peptide moiety, a nucleotide sequence encoding the polypeptide can be used in gene therapy in accordance with established principles.

METHODS

Nucleotide sequence modification methods

For example, a peptide addition may be constructed from two or more nucleotide sequences encoding a polypeptide of interest with a peptide addition, the sequences being sufficiently homologous to allow recombination between the sequences, in particular in the part thereof encoding the peptide addition. The combination of nucleotide sequences or sequence parts is conveniently conducted by methods known in the art, for instance methods which involve homologous cross-over such as disclosed in US 5,093,257, or methods which involve gene shuffling, i.e., recombination between two or more homologous nucleotide sequences resulting in new nucleotide sequences having a number of nucleotide alterations when compared to the starting nucleotide sequences. In order for homology based nucleic acid shuffling to take place the relevant parts of the nucleotide sequences are preferably at least 50% identical, such as at least 60% identical, more preferably at least 70% identical, such as at least 80% identical. The recombination can be performed in vitro or in vivo. Examples of suitable in vitro gene shuffling methods are disclosed by Stemmer et al (1994), Proc. Natl. Acad. Sci. USA; vol. 91, pp. 10747-10751; Stemmer (1994), Nature, vol. 370, pp. 389-391; Smith (1994), Nature vol. 370, pp. 324-325; Zhao et al., Nat. Biotechnol. 1998, Mar; 16(3): 258-61; Zhao H. and Arnold, FB, Nucleic Acids Research, 1997, Vol. 25. No. 6 pp. 1307-1308; Shao et al., Nucleic Acids Research 1998, Jan 15; 26(2): pp. 681-83; and WO 95/17413. Example of a suitable in vivo shuffling method is disclosed in WO 97/07205.

Furthermore, a peptide addition can be constructed by preparing a randomly mutagenized library, conveniently prepared by subjecting a nucleotide sequence encoding the polypeptide part of a conjugate of the invention to random mutagenesis to create a large number of mutated nucleotide sequences. While the random mutagenesis can be entirely random, both with respect to where in the nucleotide sequence the mutagenesis occurs and with respect to the nature of mutagenesis, it is preferably conducted so as to randomly mutate only the part of the sequence that encode the peptide addition. Also, the random mutagenesis can be directed towards introducing certain types of amino acid residues, in particular amino acid residues containing an attachment group, at random into the polypeptide molecule or at random into peptide addition part thereof. Besides substitutions, random mutagenesis can also cover random introduction of insertions or deletions. Preferably, the insertions are made in reading frame, e.g., by performing multiple introduction of three nucleotides as described by Hallet et al., Nucleic Acids Res. 1997, 25(9):1866-7 and Sondek and Shrotle, Proc Natl. Acad. Sci USA 1992, 89(8):3581-5.

The random mutagenesis (either of the whole nucleotide sequence or more preferably the part thereof encoding the peptide addition) can be performed by any suitable method. For example, the random mutagenesis is performed using a suitable physical or chemical mutagenizing

agent, a suitable oligonucleotide, PCR generated mutagenesis or any combination of these mutagenizing agentsand/or other methods according to state of the art technology, e.g. as disclosed in WO 97/07202.

Error prone PCR generated mutagenesis, e.g. as described by J.O. Deshler (1992), GATA 9(4): 103-106 and Leung et al., Technique (1989) Vol. 1, No. 1, pp. 11-15, is particularly useful for mutagenesis of longer peptide stretches (corresponding to nucleotide sequences containing more than 100 bp) or entire genes, and are preferably performed under conditions that increase the misincorporation of nucleotides.

Random mutagenesis based on doped or spiked oligonucleotides or by specific sequence oligonucleotides, is of particular use for mutagenesis of the part of the nucleotide sequence encoding the peptide addition.

Random mutagenesis of the part of the nucleotide sequence encoding the peptide addition can be performed using PCR generated mutagenesis, in which one or more suitable oligonucleotide primers flanking the area to be mutagenized are used. In addition, doping or spiking with oligonucleotides can be used to introduce mutations so as to remove or introduce attachment groups for the relevant non-peptide moiety. State of the art knowledge and computer programs (e.g. as described by Siderovski DP and Mak TW, Comput. Biol. Med. (1993) Vol. 23, No. 6, pp. 463-474 and Jensen et al. Nucleic Acids Research, 1998, Vol. 26, No. 3) can be used for calculating the most optimal nucleotide mixture for a given amino acid preference. The oligonucleotides can be incorporated into the nucleotide sequence encoding the peptide addition by any published technique using e.g. PCR, LCR or any DNA polymerase or ligase.

According to a convenient PCR method the nucleotide sequence encoding the polypeptide part of a conjugate of the invention and in particular the peptide addition thereof is used as a template and, e.g., doped or specific oligonucleotides are used as primers. In addition, cloning primers localized outside the targetted region can be used. The resulting PCR product can either directly be cloned into an appropriate expression vector or gel purified and amplified in a second PCR reaction using the cloning primers and cloned into an appropriate expression vector.

In addition to the random mutagenesis methods described herein, it is occasionally useful to employ site specific mutagenesis techniques to modify one or more selected amino acids in the peptide addition, in particular to optimise the peptide addition with respect to the number of attachment groups.

Furthermore, random elongation mutagenesis as described by Matsuura et al, *op cit* can be used to construct a nucleotide sequence encoding the polypeptide part of a conjugate of the invention having a C-terminal peptide addition. Construction of a nucleotide sequence encoding the polypeptide part of a conjugate of the invention having an N-terminal peptide addition can be constructed in an analogous way.

Also, the methods disclosed in WO 97/04079, the contents of which are incorporated herein by reference, can be used for constructing a nucleotide sequence encoding the polypeptide part of a conjugate of the invention.

The nucleotide sequence(s) or nucleotide sequence region(s) to be mutagenized is typically present on a suitable vector such as a plasmid or a bacteriophage, which as such is incubated with or otherwise exposed to the mutagenizing agent. The nucleotide sequence(s) to be mutagenized can also be present in a host cell either by being integrated into the genome of said cell or by being present on a vector harboured in the cell. Alternatively, the nucleotide sequence to be mutagenized is in isolated form. The nucleotide sequence is preferably a DNA sequence such as a cDNA, genomic DNA or synthetic DNA sequence.

Subsequent to the incubation with or exposure to the mutagenizing agent, the mutated nucleotide sequence, normally in amplified form, is expressed by culturing a suitable host cell

carrying the nucleotide sequence under conditions allowing expression to take place. The host cell used for this purpose is one, which has been transformed with the mutated nucleotide sequence(s), optionally present on a vector, or one which carried the nucleotide sequence during the mutagenesis, or any kind of gene library.

Design of peptide addition

One example of a useful guide for designing an N-terminal peptide addition containing N-glycosylation sites is characterized by the following formula:

$Y^1(NXT/S)Y^2(NXT/S)_ZY^3$-P,

wherein each of Y^1, Y^2 and Y^3 independently is absent or 1, 2, 3 or 4 amino acid residues of any type, X a single amino acid residue of any type except for proline, Z any integer between 0 and 6, T/S a threonine or serine residue, preferably a threonine residue, and N and P has the meaning defined elsewhere herein.

In a first step about 10 different muteins are made that has the above formula. For instance, the about 10 muteins are designed on the basis that each of Y^1, Y^2 and Y^3 independently is 1 or 2 alanine residues or is absent, Z any integer between 0 and 5, T/S threonine, and X alanine. Based on, e.g., *in vitro* bioactivity and half-life results obtained with these muteins (or any other relevant property), optimal number(s) of amino acids and glycosylation(s) can be determined and new muteins can be constructed based on this information. The process is repeated until an optimal glycosylated polypeptide is obtained.

Alternatively, random mutagenesis may be used for creating N-terminally extended polypeptides. For instance, a random mutagenized library is made on the basis of the above formula. Doped oligonucleotides are synthesized coding for one amino acid residue in position X (the amino acid residue being different from proline), each of Y^1, Y^2 and Y^3 independently is 0, 1 or 2 amino acid residues of any type, Z is 2 and T is threonine and used for constructing the random mutagenized library.

One example of a useful guide for designing an N-terminal peptide addition containing a PEGylation attachment group is characterized by the following formula using a lysine residue as an example of a PEGylation site. It will be understood that peptide additions with other attachment groups can be designed in an analogous way.

$Y^1(K)Y^2(K)_ZY^3$-P,

wherein each of Y^1, Y^2 and Y^3 independently is 0, 1, 2, 3 or 4 amino acid residues of any type except lysine, Z an integer between 0 and 6, K lysine, and P is as defined elsewhere herein.

In a first step about 10 different muteins are made that has the above formula. For instance, the about 10 muteins are designed on the basis that each of Y^1, Y^2 and Y^3 independently is 1 or 2 alanine residues or is absent, Z any integer between 0 and 5, and X alanine. The muteins are then PEGylated with10 kDa PEG (e.g. using mPEG-SPA). Based on, e.g., *in vitro* bioactivity and half-life results obtained with these muteins (or any other relevant property), optimal number(s) of amino acids and PEGylation sites can be determined and new muteins can be constructed based on this information. The process is repeated until an optimal PEGylated polypeptide is obtained.

Alternatively, random mutagenesis may be performed by making a random mutagenized library based on the above formula. Doped oligonucleotides are synthesized coding for one amino acid residue in position X (expect proline) and each of Y^1, Y^2 and Y^3 independently is 0, 1 or 2

amino acid residues of any type, and Z is 2 and used for constructing the random mutagenized library.

Activity Assay using PNP-glucopyranoside substrate

The enzymatic activity of recombinant glucocerebrosidase was measured using p-nitrophenyl-β-D-glucopyranoside (PNP-glucopyranoside) as a substrate. Hydrolysis of this substrate generates p-nitrophenyl, which can be quantified by measuring absorption at 405 nm using a spectrophotometer, as previously described (Friedmann et al., 1999, Blood 93; 2807-2816). The assay was carried out under conditions which partially inhibit non-glucocerebrosidase glucosidase activities, i.e., by using a phosphate/citrate buffer pH=5.5, 0.25 % Triton X-100 and 0.25 % taurocholate.

The assay was run in a final volume of 200 µl, containing 120 mM phosphate/citrate buffer, pH=5.5, 1 mM EDTA, pH=8.0, 0.25 % Triton X-100, 0.25 % taurocholate, 4 mM β-mercaptoethanol and 4 mM PNP-glucopyranoside. The enzymatic hydrolysis was initiated by adding glucocerebrosidase and the reaction was allowed to proceed for 1 hour at 37°C before stopping the reaction by adding 50 µl 1 M NaOH and measuring absorption at 405 nm. A reference standard curve of p-nitrophenyl, assayed in parallel, was used to quantify concentrations of glucocerebrosidase in samples to be tested.

EXAMPLES

EXAMPLE 1

PRODUCTION OF GLUCOCEREBROSIDASE

Cloning and Expression in Yeast Cells

A synthetic glucocerebrosidase gene encoding the enzyme was designed with the Mat-alpha prepro sequence from *S. cerevisiae* added in front of the enzyme in order to facilitate secretion of the enzyme from the *S. cerevisiae* or *Pichia pastoris* cells. The codon usage of the human glucocerebrosidase gene was changed to codons optimal for expression in *S. cerevisiae* and restriction enzyme sites were added in order to make unique cloning sites available (see the nucleotide sequence shown in Fig. 1). The synthetic glucocerebrosidase gene was prepared by obtaining oligonucleotides with 20-25 bp overlap covering the whole gene (see the nucleotide sequence of the oligonucleotides shown in Fig. 2). The oligonucleotides were assembled in a PCR reaction using the Pfx polymerase (Life Technologies) under the recommended conditions supplied with the Pfx polymerase with an annealing temperature of 45°C. 2, 5 and 10 µl of the resulting PCR product was used in a second PCR using a primer covering the 5'-end and another anti-sense primer covering the 3'-end of the synthetic gene. The resulting PCR reaction giving the largest amount of PCR product of the expected length was agarose gel purified. Following agarose purification the PCR fragment was digested with BamHI and XbaI and cloned into BamHI and XbaI sites of the pJSO37 expression vector (Okkels, Ann. New York Acad. Sci. 782, 202-207, 1996). Several clones were sequenced and one of those without any PCR generated missense mutations was named pSV1 and transformed into the *S. cerevisiae* strain YNG318 (available from the American Type Culture Collection 10801 University Boulevard, Manassas, VA 20110-2209, USA as ATCC 208973 and described by Rourke et al. J. Biol. Chem. 272, pp. 9720-9727, 1997). Expression of the resulting *S. cerevisiae* transformants was done as described in Okkels, Ann. New York Acad. Sci. 782, 202-207, 1996.

Cloning and Expression in Insect Cells

A human fibroblast cDNA library was obtained from Clontech (Human fibroblast skin cDNA cloned in lambda-gt11, cat# HL1052b). Lambda DNA was prepared from the library by standard methods and used as a template in a PCR reaction with either SO49 and SO50 as primers (amplify the GCB coding region with the human signal peptide from the second ATG) or SO50 and SO51 as primers (amplify the mature part of the glucocerebrosidase coding region) (see Fig. 3).

The PCR products were reamplified with the same primers and agarose gel purified. Subsequently the SO49/50 PCR product was digested with BglII and EcoRI and cloned into the pBlueBac 4.5 vector (Invitrogen, Carlsbad, CA, USA, Carlsbad, CA, USA) digested with BamHI and EcoRI. The resulting plasmid was used for infection of insect cells with the glucocerebrosidase being partly secreted from the cells due to the human signal sequence as described in Martin et al., DNA 7, pp. 99-106, 1988. The SO50/51 PCR product was digested with SacI and EcoRI and cloned into the pBlueBac 4.5 vector (Invitrogen, Carlsbad, CA, USA) digested with the same enzymes resulting in the pGCBmat plasmid. Two different signal sequences were inserted upstream of the mature glucocerebrosidase codons in order to increase the secreted amount of enzyme. The baculovirus ecdysteroid UDPglucosyltransferase (egt) signal sequence (Murphy et al., Protein Expression and Purification 4, 349-357, 1993) was inserted by annealling SO52 and SO53 (Fig. 3) and the human pancreatic lipase signal sequence (Lowe et al., J. Biol. Chem. 264, 20042, 1989) was inserted by annealling SO54 and SO55 (Fig. 3) and cloning them into the NheI and SacI digested pGCBmat plasmid. Infection of *Spodoptera frugiperda* (Sf9) cells of the resulting plasmid was done according to the protocols from Invitrogen, Carlsbad, CA, USA.

Purification of glucocerebrosidase wildtype and muteins produced in insect cells

Polypeptides with glucocerebrosidase activity were purified as described in US 5,236,838, with minor modifications. Cells were removed from the culture medium by centrifugation (10 min at 4000 rpm in a Sorvall RC5C centrifuge) and the supernatant microfiltrated using a 0.22 µm filter prior to purification. Before application on the first Hydrophobic Interaction Chromatography (HIC) capture step DTT was added to 1 mM and the culture supernatant diluted with distilled water to obtain a low ionic strength (≤ 8 mS/cm at room temperature). Under these conditions at pH 6 (or lower) GC still binds to a Toyopearl butyl resin (TosoHaas) equilibrated in 50 mM sodium citrate, 20 % (v/v) ethylene glycol, 1 mM DTT, pH 5.0 (buffer A). The binding capacity of a 70 ml (2.6 x 13 cm) column is sufficient for at least 1000 ml of culture medium. After application, the resin used at the capture step was washed with at least 3 column volumes of Buffer A (until the absorbance at 280 nm reaches baseline level) and GC was eluted with a linear gradient over 60 min from 0% to 100% buffer B (50 mM sodium citrate, 80% (v/v) ethylene glycol, 1 mM DTT, pH 5.0) at a flow rate of 1 ml/min. Fractions were collected and assayed for glucocerebrosidase activity using the PNP-glucopyranoside assay. Usually, wildtype glucocerebrosidase starts to elute at approx. 70% (v/v) ethylene glycol. Glucocerebrosidase enriched fractions from the first process step were pooled and diluted approx. 4 times with a buffer containing 50 mM sodium citrate, 1 mM DTT, pH 5.0 to reduce the ethylene glycol content to 20% (or lower). In the second HIC purification step the diluted and partially purified glucocerebrosidase was applied on a Toyopearl phenyl resin (TosoHaas). The employed column of 45 ml (2.6 x 8 cm) was equilibrated in 50 mM sodium citrate, 1 mM DTT, pH 5.0 (Buffer A) before use.

After application, the resin was washed with at least 3 column volumes of 50 mM sodium citrate, pH 5 (until the absorbance at 280 nm reaches baseline level) and glucocerebrosidase was then eluted with a linear ethanol gradient from 0% to 100% buffer B (50 mM sodium citrate, 50% (v/v)

ethanol, 1 mM DTT, pH 5.0) over 60 min at a flow rate of 1 ml/min. Highly purified fractions of glucocerebrosidase (wildtype ≥ 95% pure), identified using the enzyme activity assay, will start to elute at approx. 40% ethanol. The purified glucocerebrosidase bulk product was dialyzed against either 50 mM sodium citrate, 1 mM DTT, pH 5.0 or 50 mM sodium citrate, 80% (v/v) ethylene glycol, 1 mM DTT, pH 5.0 to retain the enzyme activity upon subsequent storage. The purified glucocerebrosidase was stored at 4 – 8 °C.

Preparation of glucocerebrosidase with N-terminal peptide addition

Nucleotide sequences encoding the following N-terminal peptide additions were added to the nucleotide sequence shown in SEQ ID NO 1 encoding glucocerebrosidase: (A-4)+(N-3)+(I-2)+(T-1) (representing an extension to the N-terminal of the amino acid sequence encoding by the nucleotide sequence shown in Fig. 1 with the amino acid residues ANIT), and (A-7)+(S-6)+(P-5)+(I-4)+(N-3)+(A-2)+(T-1) (ASPINAT).

A nucleotide sequence encoding the N-terminal peptide addition (A-4)+(N-3)+(I-2)+(T-1) was prepared by PCR using the following conditions:

PCR 1:
Template: 10 ng pBlueBac5 with wt GCB,
primer SO60: 5'-CAGCTGGCCATGGGTACCCGG-3' and
primer SO85: 5'-TGGGCATCAGGTGCCAACATTACAGCCCGCCCCTGCATCCCTAAAAGC-3'
BIO-X-ACT™ DNA polymerase (Bioline, London, U.K.)
1xOptiBuffer™ (Bioline, London, U.K.)
30 cycles of 96°C 30s, 55°C 30s, 72°C 1 min
PCR 2:
Template: 10 ng pBlueBac5 with wt GCB,
Baculo virus forward primer: 5'-TTTACTGTTTTCGTAACAGTTTTG-3' and
primer SO86: 5'- GCAGGGGCGGGCTGTAATGTTGGCACCTGATGCCCACGACACTGCCTG-3'
BIO-X-ACT™ DNA polymerase (Bioline, London, U.K.)
1xOptiBuffer™ (Bioline, London, U.K.)
30 cycles of 96°C 30s, 55°C 30s, 72°C 1 min
PCR 3:
3 µl of agarose gel purified PCR1 and PCR2 products (app. 10 ng)
Baculo virus forward primer: 5'-TTTACTGTTTTCGTAACAGTTTTG-3'
primer SO60: 5'-CAGCTGGCCATGGGTACCCGG-3'
BIO-X-ACT™ DNA polymerase (Bioline, London, U.K.)
1xOptiBuffer™ (Bioline, London, U.K.)
30 cycles of 96°C 30s, 55°C 30s, 72°C 1 min

PCR 3 was agarose gel purified and digested with NheI and NcoI and cloned into pBluebac4.5+wtGCB digested with NheI and NcoI.

After confirmation of the correct mutations by DNA sequencing the plasmid was transfected into insect cells using the Bac-N-Blue™ transfection kit from Invitrogen, Carlsbad, CA, USA. Expression of the muteins was tested by western blotting and by activity measurement of the muteins using the above described activity assay.

Enzymatic activity in the PNP assay of wildtype glucocerebrosidase (SEQ ID NO 1) expressed in the expression vector pVL1392 (Pharmingen, USA) in insect cells (Sf9) using an analogous method to that described in EXAMPLE 1 above gave 13 units/L, while the N-terminal peptide addition ASPINAT gave 28.5 units/L.

When subjecting these insect cell expressed N-terminally extended glycosylated polypeptides to N-terminal amino acid sequence analysis (using Procize from PE Biosystems, Foster City, CA), the sequencing cycle was blank for the Asn residue in both ANIT and ASPINAT N-terminal peptide additions, demonstrating that the introduced glycosylation site is glycosylated.

When subjecting conjugate with ASPINAT addition to mass spectrophometry using the MALDI-TOF techniques on the Voyager DERP instrument (from PE-Biosystems, Foster City, CA) the following results were obtained:

The wildtype and ASPINAT-extended wildtype expressed in insect cells gave average masses very close to the calculated mass of 59,727 Da and 61,421 Da, respectively, assuming that four glycosylation sites were occupied by the carbohydrates $FucGlcNAc_2Man_3$.

The thermostability of wildtype with and without peptide addition is shown in Fig. 5. The thermostability was determined in the PNP assay using 10 mU of enzyme per time point. After incubation for 0-180 minutes at either pH 5.5 or pH 7.4 at 37 C the enzyme activity was determined.

CLAIMS

1. A conjugate of a polypeptide and a non-peptide moiety, wherein the polypeptide part of the conjugate comprises the primary structure,
NH_2- X-P-COOH or NH_2-P-X-COOH,
wherein
X is a peptide addition comprising or contributing to an attachment group for the non-peptide moiety, and
P is a polypeptide of interest.

2. A conjugate of a polypeptide and a non-peptide moiety, wherein the polypeptide part of the conjugate comprises the primary structure NH_2-P_x-X-P_y-COOH, wherein
P_x is an N-terminal part of a polypeptide P of interest,
P_y is a C-terminal part of said polypeptide P, and
X is a peptide addition comprising or contributing to an attachment group for the non-peptide moiety.

3. The conjugate according to claim 1 or 2, wherein P is a mature polypeptide.

4. The conjugate according to claim 2 or 3, wherein P_x is a non-structural N-terminal part of a mature polypeptide P, and P_y is a structural C-terminal part of said mature polypeptide, or P_x is a structural N-terminal part of a mature polypeptide P, and P_y is a non-structural C-terminal part of said mature polypeptide.

5. The conjugate according to any of claims 1-4, wherein P is a native polypeptide.

6. The conjugate according to any of claims 1-4, wherein P is a variant of a native polypeptide.

7. The conjugate according to claim 6, wherein P comprises at least one introduced and/or at least one removed attachment group for the non-peptide moiety as compared to the corresponding native polypeptide.

8. The conjugate according to any of claims 1-7, wherein P is of mammalian origin.

9. The conjugate according to claim 8, wherein P is of human origin.

10. The conjugate according to any of claims 1-9, wherein P is a therapeutic polypeptide.

11. The conjugate according to any of claims 1-10, wherein P is selected from the group consisting of an antibody or antibody fragment, a plasma protein, an erythrocyte or thrombocyte protein, a cytokine, a growth factor, a profibrinolytic protein, a protease inhibitor, an antigen, an enzyme, a ligand, a receptor, or a hormone.

12. The conjugate according to claim 11, wherein P is an enzyme having a therapeutical effect on patients with a lysosomal storage disease.

13. The conjugate according to claim 12, wherein P is selected from the group consisting of glucocerebrosidase, α-L-iduronidase, acid α-glucosidase, α-galactosidase, acid sphingomyelinase, and hexosaminidase.

14. The conjugate according to claim 12 or 13, wherein the non-peptide moiety is a sugar moiety and the attachment group is an N-glycosylation site.

15. The conjugate according to any of claims 1-7 or 10-14, wherein P is of microbial origin.

16. The conjugate according to claim 15, wherein P is a microbial enzyme.

17. The conjugate according to claim 16, wherein P is selected from the group consisting of protease, amylase, amyloglucosidase, pectinase, lipase and cutinase.

18. The conjugate according to any of claims 1-17, wherein X comprises 1-500 amino acid residues.

19. The conjugate according to claim 18, wherein X comprises 2-50 amino acid residues, such as 3-20 amino acid residues.

20. The conjugate according to any of claims 1-19, wherein X comprises 1-20, in particular 1-10 attachment groups for the non-peptide moiety.

21. The conjugate according to any of the preceding claims, wherein X comprises at least one attachment group within a stretch of 30 amino acid residues, such as at least one within 20 amino acid residues, in particular at least one within 10 amino acid residues, in particular 1-3 attachment groups.

22. The conjugate according to any of claims 1-21, wherein X comprises at least two attachment groups for the non-peptide moiety, wherein two of said amino acid residues are separated by at most 10 amino acid residues, none of which comprises an attachment group for the non-peptide moiety.

23. The conjugate according to any of claims 7-22, wherein the polypeptide P comprises at least one introduced attachment group for the non-peptide moiety, in particular 1-5 introduced attachment groups.

24. The conjugate according to any of claims 7-23, wherein the polypeptide P comprises at least one removed attachment group for the non-peptide moiety, in particular 1-5 removed attachment groups.

25. The conjugate according to any of claims 1-24, wherein the non-peptide moiety is a sugar moiety and the attachment group is an *in vivo* glycosylation site.

26. The conjugate according to claim 14 or 25, wherein X has the structure X_1-N-X_2-T/S-Z, wherein X_1 is a peptide comprising at least one amino acid residue or is absent, X_2 is any amino acid residue different from P, and Z is absent or a peptide comprising at least one amino acid residue, the N-terminal amino acid residue of which is different from P.

27. The conjugate according to claim 26, wherein X_1 is absent, X_2 is an amino acid residue selected from the group consisting of I, A, G, V and S, and Z comprises at least one amino acid residue, the N-terminal amino acid residue of which is different from P.

28. The conjugate according to claim 27, wherein Z is a peptide comprising 1-50 amino acid residues, preferably comprising 1-10 glycosylation sites.

29. The conjugate according to claim 26, wherein X_1 comprises at least one amino acid residue, X_2 is an amino acid residue selected from the group consisting of I, A, G, V and S, and Z is absent.

30. The conjugate according to claim 29, wherein X_1 is a peptide comprising 1-50 amino acid residues, preferably comprising 1-10 glycosylation sites.

31. The conjugate according to claim 26-30, wherein X comprises a peptide sequence selected from the group consisting of INAT/S, GNIT/S, VNIT/S, SNIT/S, ASNIT/S, NIT/S, SPINAT/S, ASPINAT/S, ANIT/SANIT/SANI, and ANIT/SGSNIT/SGSNIT/S, wherein T/S is either a T or an S residue, preferably a T residue.

32. The conjugate according to any of claims 26-31, wherein X is selected from the group consisting of INAT/S, GNIT/S, VNIT/S, SNIT/S, ASNIT/S, NIT/S, SPINAT/S, ASPINAT/S, ANIT/SANIT/SANI, and ANIT/SGSNIT/SGSNIT/S, wherein T/S is either a T or an S residue, preferably a T residue.

33. The conjugate according to any of claims 25-33, wherein the polypeptide P is a variant of a native polypeptide which, as compared to said native polypeptide, comprises at least one introduced or at least one removed glycosylation site.

34. The conjugate according to claim 33, wherein the polypeptide P comprises at least one introduced glycosylation site, in particular 1-5 introduced glycosylation sites.

35. The conjugate according to claim 33 or 34, wherein the glycosylation site is introduced so that the N residue of said glycosylation site is exposed at the surface of the polypeptide, when folded in its active form.

36. The conjugate according to any of claims 25-35, wherein X has an N residue in position –2 or –1, and P has a T or an S residue in position +1 or +2, respectively, the residue numbering being made relative to the N-terminal amino acid residue of P.

37. The conjugate according to any of claims 25-36, which further comprises at least one polymer molecule.

38. The conjugate according to any of claims 1-13 or 15-22, wherein the non-peptide moiety is selected from the group consisting of a polymer molecule, a lipophilic group and an organic derivatizing agent.

39. The conjugate according to claim 38, wherein the attachment group for the non-peptide moiety is one present on an amino acid residue selected from the group consisting of the N-terminal amino acid residue of the polypeptide part of the conjugate, the C-terminal residue of the polypeptide part of the conjugate, lysine, cysteine, arginine, glutamine, aspartic acid, glutamic acid, serine, tyrosine, histidine, phenylalanine and tryptophan.

40. The conjugate according to claim 39, wherein the attachment group for the non-peptide moiety is an ε-amino group.

41. The conjugate according to any of claims 38-40, wherein the X comprises at least two attachment groups for the non-peptide moiety.

42. The conjugate according to any of claims 38-41, wherein the polypeptide P is a variant of a native polypeptide, which as compared to said native polypeptide, comprises at least one introduced and/or at least one removed attachment group for the non-peptide moiety.

43. The conjugate according to claim 42, wherein the polypeptide P comprises at least one introduced attachment group, in particular 1-5 introduced attachment groups.

44. The conjugate according to any of claims 38-43, the polypeptide of which is glycosylated.

45. The conjugate according to claim 44, wherein X further comprises at least one glycosylation site.

46. The conjugate according to any of the preceding claims, which comprises at two different non-peptide moieties, in particular a sugar moiety and a polymer molecule.

47. The conjugate according to any of the preceding claims, which has a molecular weight of at least 67 kDa, in particular at least 70 kDa.

48. The conjugate according to any of the preceeding claims, which has at least one of the following properties relative to the polypeptide P, the properties being measured under comparable conditions:
in vitro bioactivity which is at least 25% of that of the polypeptide P as measured under comparable conditions, increased affinity for a mannose receptor or other carbohydrate receptors, increased serum half-life, increased functional *in vivo* half-life, reduced renal clearance, reduced immunogenicity, increased resistance to proteolytic cleavage, improved targeting to lysosomes, macrophages and/or other subpopulations of human cells, improved stability in production, improved shelf life, improved formulation, e.g. liquid formulation, improved purification, improved solubility, and/or improved expression.

49. The conjugate according to any of claims 1-14 or 17-48, which is a glycosylated polypeptide with the primary structure NH_2-X-P-COOH, wherein P is a mature, enzymatically active form of glucocerebrosidase and X is as defined in claim 31.

50. A nucleotide sequence encoding the polypeptide part of the conjugate according to any of claims 1-49.

51. A vector comprising the nucleotide sequence according to claim 50.

52. A host cell transformed or transfected with a nucleotide sequence according to claim 50, or a vector according to claim 51.

53. The host cell according to claim 52, which is a glycosylating host cell.

54. The host cell according to claim 53, which is a mammalian cell, an invertebrate cell such as an insect cell, a yeast cell or a plant cell, or a transgenic animal.

55. The host cell according to claim 54, wherein the yeast cell is selected from the group consisting of *Saccharomyces cerevisiae, Pichia pastoris, Hansenula* spp, the insect cell is an Sf9 cell or a Hi5 cell, and the mammalian cell is selected from the group consisting of CHO, BHK and COS cells.

56. A method of producing the polypeptide part of the conjugate according to any of claims 1-49, comprising culturing a host cells according to any of claims 52-55 under conditions permitting expression of the polypeptide and recovering the polypeptide from the culture.

57. A method of producing a conjugate according to any of claim 1-49 comprising a glycosylated polypeptide, comprising culturing a glycosylating host cell according to any of claims 52-55 under conditions permitting the expression of the glycosylated polypeptide and recovering the polypeptide from the culture.

58. A method of producing a conjugate according to any of claims 1-49 comprising a non-peptide moiety selected from the group consisting of a polymer molecule, a lipophilic group and an organic derivatizing agent, which method comprises subjected the polypeptide part of the conjugate to conjugation to the non-peptide moiety under conditions for the conjugation to take place.

59. The method according to claim 58, wherein the polypeptide part of the conjugate is prepared by the method according to 56 or 57.

60. A method of preparing a nucleotide sequence according to claim 50, which method comprises

a) subjecting a nucleotide sequence encoding the polypeptide P to elongation mutagenesis,
b) expressing the mutated nucleotide sequence obtained in step a),
c) conjugating polypeptides expressed in step b) to the non-peptide moiety to be used for preparing the relevant polypeptide conjugate,
d) selecting conjugates comprising at least one non-peptide moiety attached to the peptide addition part of the polypeptide, and
e) isolating a nucleotide sequence encoding the polypeptide part of conjugates selected in step d).

61. The method according to claim 60, which further comprises screening conjugates resulting from step c) for at least one improved property prior, and wherein the selection step d) further comprises selecting conjugates having such improved property.

62. The method according to claim 60 or 61, wherein the elongation mutagenesis is conducted so as to enrich for codons encoding an amino acid residue comprising an attachment group for the non-peptide moeity.

63. The method according to claim 60 or 61, wherein the elongation mutagenesis is conducted so as to enrich for codons required for introduction of an *in vivo* glycosylation site.

64. The method according to any of claims 60-62, which further comprises subjecting the part of the nucleotide sequence encoding P to mutagenesis to remove and/or introduce amino acid residues comprising attachment groups for the non-peptide moiety.

65. The method according to any of claims 60-64, wherein the selection in step c) is performed so as to select a conjugate having at least one of the properties defined in claim 48.

66. A method of producing a glycosylated polypeptide encoded by a nucleotide sequence prepared according to claims 60-65, wherein the nucleotide sequence encoding the polypeptide selected in step c) is expressed in a glycosylating host cell and the resulting glycosylated expressed polypeptide is recovered.

67. A method of improving one or more selected properties of a polypeptide P of interest, which method comprises

a) preparing a nucleotide sequence encoding a polypeptide with the primary structure

NH_2-X-P-COOH or NH_2-P-X-COOH,

wherein
X is a peptide addition comprising or contributing to an attachment group for a non-peptide moiety that is capable of conferring the selected improved property/ies to the polypeptide P, when conjugated thereto,
b) expressing the nucleotide sequence of a) in an suitable host cell,
c) conjugating the expressed polypeptide of b) to the non-peptide moiety, and
d) recovering the conjugate resulting from step c).

68. The method according to claim 67, wherein the polypeptide P is as defined in any of claims 1-49.

69. The method according to claim 68, wherein the non-peptide moiety is a sugar moiety, the host cell in step b) is a glycosylating host cell, and the conjugation in step c) is achieved by *in vivo* glycosylation during the expression step b).

70. The method according to any of claims 67-69, wherein the nucleotide sequence of step a) is prepared by subjecting a nucleotide sequence encoding the polypeptide P to random elongation mutagenesis.

71. The method according to claim 70, wherein the random elongation mutagenesis is conducted so as to enrich for codons encoding an aminio acid residue comprising an attachment group for the non-peptide moiety, in particular an *in vivo* glycosylation site.

72. The method according to any of claims 67-71, wherein, in the preparation of the nucleotide sequence of a), the part of the nucleotide sequence encoding the polypeptide P is subjected to mutagenesis to remove and/or introduce an attachment group for the non-peptide moiety.

73. The method according to any of claims 66-72, wherein the property/ies to be improved is/are selected from the properties defined in claim 48.

SEQ ID NO: 1
Amino acid sequence of wt glucocerebrosidase
1 ARPCIPKSFG YSSVVCVCNA TYCDSFDPPT FPALGTFSRY ESTRSGRRME
LSMGPIQANH
61 TGTGLLLTLQ PEQKFQKVKG FGGAMTDAAA LNILALSPPA QNLLLKSYFS
EEGIGYNIIR
121 VPMASCDFSI RTYTYADTPD DFQLHNFSLP EEDTKLKIPL IHRALQLAQR
PVSLLASPWT
181 SPTWLKTNGA VNGKGSLKGQ PGDIYHQTWA RYFVKFLDAY AEHKLQFWAV
TAENEPSAGL
241 LSGYPFQCLG FTPEHQRDFI ARDLGPTLAN STHHNVRLLM LDDQRLLLPH
WAKVVLTDPE
301 AAKYVHGIAV HWYLDFLAPA KATLGETHRL FPNTMLFASE ACVGSKFWEQ
SVRLGSWDRG
361 MQYSHSIITN LLYHVVGWTD WNLALNPEGG PNWVRNFVDS PIIVDITKDT
FYKQPMFYHL
421 GHFSKFIPEG SQRVGLVASQ KNDLDAVALM HPDGSAVVVV LNRSSKDVPL
TIKDPAVGFL
481 ETISPGYSIH TYLWRRQ

Figure 1:
The nucleotide sequence of the synthetic GCB gene with a His-tag designed for expression in S. cerevisiae. The introduced unique restriction enzyme sites are underlined.

CGGGGATCCGAATTCAACATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGC
ATCCTCCGCATTAGCTGCACCGGTCAACACTACAACAGAAGATGAAACGGCACAAATT
CCGGCTGAAGCTGTCATCGGTTACTTAGATTTAGAAGGGGATTTCGATGTTGCTGTTTT
GCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCA
TTGCTGCTAAAGAAGAAGGGGTCTCGAGAGATAAAAGACAAAAGCACCAACACCAAC
ATCAACATCAACATCAACACCAAGCGCGCCCCATGTATTCCTAAGTCTTTCGGTTACTCT
TCCGTTGTTTGTGTCTGTAATGCCACATACTGTGACTCCTTTGACCCACCGACCTTTCCT
GCTTTGGGTACCTTCTCCAGATATGAATCTACTCGTTCCGGCCGTAGAATGGAATTGAG
TATGGGTCCAATTCAAGCTAATCACACTGGCACTGGTCTTCTACTGACCTTGCAACCAG
AACAAAAGTTCCAAAAAGTCAAGGGATTTGGTGGCGCCATGACAGATGCTGCCGCTCT
GAACATCCTTGCCTTGTCACCACCAGCCCAAAATTTGCTATTGAAATCTTACTTCTCTG
AAGAAGGAATCGGTTATAACATTATCCGTGTTCCTATGGCCTCTTGTGACTTCTCCATC
AGAACCTACACTTATGCAGACACCCCTGATGATTTCCAATTGCACAACTTCTCTTTGCC
AGAGGAAGATACCAAGTTGAAGATTCCCCTGATTCACCGTGCTCTACAGTTGGCCCAA
AGACCAGTTTCCCTACTTGCTTCTCCTTGGACTTCCCCTACCTGGTTAAAGACTAATGG
CGCTGTTAATGGTAAGGGTTCTCTCAAGGGACAGCCAGGAGACATCTACCACCAAACC
TGGGCCAGATACTTTGTTAAGTTCTTGGATGCCTATGCTGAACACAAGTTACAATTCTG
GGCAGTCACTGCTGAAAATGAACCTTCTGCTGGTCTGTTGTCTGGTTACCCATTCCAAT
GCTTGGGCTTCACCCCTGAACATCAAAGAGACTTCATTGCCAGAGATCTAGGTCCTACC
TTGGCCAACTCCACTCACCACAATGTCAGACTATTGATGCTGGATGACCAAAGGTTGCT
GCTACCACACTGGGCAAAGGTGGTTTTGACTGACCCAGAAGCTGCTAAATATGTTCAT
GGCATTGCTGTCCATTGGTACTTGGACTTTTTGGCTCCAGCCAAAGCCACCTTAGGCGA
AACTCACAGATTATTCCCCAACACCATGTTGTTTGCTTCAGAAGCATGCGTTGGCTCCA
AGTTCTGGGAACAAAGTGTTAGACTAGGCTCCTGGGATAGAGGTATGCAATACTCTCA
CTCTATCATCACTAACTTATTGTACCATGTCGTCGGCTGGACCGACTGGAACCTTGCCC

TGAACCCAGAAGGAGGTCCTAATTGGGTTCGTAACTTTGTCGACAGTCCAATCATTGTT
GACATCACCAAGGACACTTTTTACAAACAACCAATGTTCTACCACTTGGGTCATTTCTC
TAAGTTCATTCCTGAAGGCTCCCAAAGAGTGGGACTAGTTGCCTCTCAAAAGAACGAC
TTGGACGCAGTTGCTTTGATGCACCCAGATGGCTCTGCTGTTGTGGTCGTTCTAAACCG
TTCCTCTAAGGATGTTCCTCTTACCATCAAGGACCCAGCTGTTGGTTTCTTGGAAACAA
TTTCACCTGGCTACTCCATTCACACCTACTTGTGGCGTAGACAATAATACCGCGGTCTA
GAGCC

Figure 2:
Oligonucleotides used for assembling the synthetic GCB gene with a His-tag (primer SO3 to SO40) and without a His-tag (SO3 to SO8 and SO11 to SO42):

SO3: 5'-CGGGGATCCGAATTCAACATGAG-3'
SO4: 5'-GGCTCTAGACCGCGGTATTATTGTC-3'
SO5:
5'-
CGGGGATCCGAATTCAACATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGC
ATCCTCCGCATTAGC-3'
SO6:
5'-
GACAGCTTCAGCCGGAATTTGTGCCGTTTCATCTTCTGTTGTAGTGTTGACCGGTGCAG
CTAATGCGGAGGATGCTGCG-3'
SO7:
5'-
CAAATTCCGGCTGAAGCTGTCATCGGTTACTTAGATTTAGAAGGGGATTTCGATGTTGC
TGTTTTGCCATTTTCC-3'
SO8:
5'-
CAGCAATGCTGGCAATAGTAGTATTTATAAACAATAACCCGTTATTTGTGCTGTTGGAA
AATGGCAAAACAGCAAC-3'
SO9:
5'-
CTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTCTCGAGAGATAAAAGACAAAA
GCACCAACACCAACATC-3'
SO10:
5'-
AGAGTAACCGAAAGACTTAGGAATACATGGGCGCGCTTGGTGTTGATGTTGATGTTGA
TGTTGGTGTTGGTGCTTTTG-3'
SO11:
5'-
CCTAAGTCTTTCGGTTACTCTTCCGTTGTTTGTGTCTGTAATGCCACATACTGTGACTCC
TTTGACCCACCG-3'
SO12:
5'-
GGAACGAGTAGATTCATATCTGGAGAAGGTACCCAAAGCAGGAAAGGTCGGTGGGTC
AAAGGAGTCACAG-3'
SO13:

5'-
CAGATATGAATCTACTCGTTCCGGCCGTAGAATGGAATTGAGTATGGGTCCAATTCAA
GCTAATCACACTGGC-3'
SO14:
5'-
GACTTTTTGGAACTTTTGTTCTGGTTGCAAGGTCAGTAGAAGACCAGTGCCAGTGTGAT
TAGCTTGAATTGG-3'
SO15:
5'-
CCAGAACAAAAGTTCCAAAAAGTCAAGGGATTTGGTGGCGCCATGACAGATGCTGCCG
CTCTGAACATCC-3'
SO16:
5'-
GAAGTAAGATTTCAATAGCAAATTTTGGGCTGGTGGTGACAAGGCAAGGATGTTCAGA
GCGGCAGCATC-3'
SO17:
5'-
AATTTGCTATTGAAATCTTACTTCTCTGAAGAAGGAATCGGTTATAACATTATCCGTGT
TCCTATGGCCTC-3'
SO18:
5'-
CATCAGGGGTGTCTGCATAAGTGTAGGTTCTGATGGAGAAGTCACAAGAGGCCATAGG
AACACGGATAATG-3'
SO19:
5'-
CACTTATGCAGACACCCCTGATGATTTCCAATTGCACAACTTCTCTTTGCCAGAGGAAG
ATACCAAGTTG-3'
SO20:
5'-
GGAAACTGGTCTTTGGGCCAACTGTAGAGCACGGTGAATCAGGGGAATCTTCAACTTG
GTATCTTCCTCTGGC-3'
SO21:
5'-
CAGTTGGCCCAAAGACCAGTTTCCCTACTTGCTTCTCCTTGGACTTCCCCTACCTGGTTA
AAGACTAATGGC-3'
SO22:
5'-
GGTAGATGTCTCCTGGCTGTCCCTTGAGAGAACCCTTACCATTAACAGCGCCATTAGTC
TTTAACCAGGTAG-3'
SO23:
5'-
GGGACAGCCAGGAGACATCTACCACCAAACCTGGGCCAGATACTTTGTTAAGTTCTTG
GATGCCTATGCTG-3'
SO24:
5'-
GCAGAAGGTTCATTTTCAGCAGTGACTGCCCAGAATTGTAACTTGTGTTCAGCATAGGC
ATCCAAGAACTTAAC-3'
SO25:

5'-
CTGCTGAAAATGAACCTTCTGCTGGTCTGTTGTCTGGTTACCCATTCCAATGCTTGGGC
TTCACCCCTGAAC-3'
SO26:
5'-
GAGTTGGCCAAGGTAGGACCTAGATCTCTGGCAATGAAGTCTCTTTGATGTTCAGGGG
TGAAGCCCAAGCATTG-3'
SO27:
5'-
CTAGGTCCTACCTTGGCCAACTCCACTCACCACAATGTCAGACTATTGATGCTGGATGA
CCAAAGGTTGCTGC-3'
SO28:
5'-
CATATTTAGCAGCTTCTGGGTCAGTCAAAACCACCTTTGCCCAGTGTGGTAGCAGCAAC
CTTTGGTCATCCAGC-3'
SO29:
5'-
CTGACCCAGAAGCTGCTAAATATGTTCATGGCATTGCTGTCCATTGGTACTTGGACTTT
TTGGCTCCAGCCAAAGC-3'
SO30:
5'-
GAAGCAAACAACATGGTGTTGGGGAATAATCTGTGAGTTTCGCCTAAGGTGGCTTTGG
CTGGAGCCAAAAAGTCC-3'
SO31:
5'-
CCCAACACCATGTTGTTTGCTTCAGAAGCATGCGTTGGCTCCAAGTTCTGGGAACAAAG
TGTTAGACTAGGC-3'
SO32:
5'-
CAATAAGTTAGTGATGATAGAGTGAGAGTATTGCATACCTCTATCCCAGGAGCCTAGT
CTAACACTTTGTTCC-3'
SO33:
5'-
CTCTATCATCACTAACTTATTGTACCATGTCGTCGGCTGGACCGACTGGAACCTTGCCC
TGAACCCAGAAGG-3'
SO34:
5'-
GTCAACAATGATTGGACTGTCGACAAAGTTACGAACCCAATTAGGACCTCCTTCTGGG
TTCAGGGCAAGG-3'
SO35:
5'-
CGACAGTCCAATCATTGTTGACATCACCAAGGACACTTTTTACAAACAACCAATGTTCT
ACCACTTGGGTC-3'
SO36:
5'-
GGCAACTAGTCCCACTCTTTGGGAGCCTTCAGGAATGAACTTAGAGAAATGACCCAAG
TGGTAGAACATTGG-3'
SO37:

5'-
CCAAAGAGTGGGACTAGTTGCCTCTCAAAAGAACGACTTGGACGCAGTTGCTTTGATG
CACCCAGATGGCTC-3'
SO38:
5'-
GGTAAGAGGAACATCCTTAGAGGAACGGTTTAGAACGACCACAACAGCAGAGCCATC
TGGGTGCATCAAAGC-3'
SO39:
5'-
CCTCTAAGGATGTTCCTCTTACCATCAAGGACCCAGCTGTTGGTTTCTTGGAAACAATT
TCACCTGGCTACTC-3'
SO40:
5'-
GGCTCTAGACCGCGGTATTATTGTCTACGCCACAAGTAGGTGTGAATGGAGTAGCCAG
GTGAAATTGTTTCC-3'
SO41 (-HIS):
5'-
CTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTCTCGAGAGATAAGCGCGCTAG
ACC-3'
SO42 (-HIS):
5'-AGAGTAACCGAAAGACTTAGGAATACATGGTCTAGCGCGCTTATCTCTCGAGACC-3'

Figure 3:
Sequence of primers used for cloning the wt GCB coding region and inserting signal peptides into the pGCBmat plasmid.

SO49 (WT-sp-BglII):
5'-CGCAGATCTGATGGCTGGCAGCCTCACAGGATTGC-3'
SO50 (WT-stop-EcoRI):
5'-CCGGAATTCCCATCACTGGCGACGCCACAGGTAGGTG-3'
SO51 (WT-mature-SacI):
5'-ACGCGAGCTCGCCCCTGCATCCCTAAAAGCTTCGG-3'
SO52 (SPegt-NheI/SacI-as)
5'-GCGTTGACGGCAGTCAGAGTTGACAGAAGGGCCAGCCAGCAAAGGATAGTCATG-3'
SO53 (SPegt-NheI/SacI-s)
5'-
CTAGCATGACTATCCTTTGCTGGCTGGCCCTTCTGTCAACTCTGACTGCCGTCAACGCA
GCT-3'
SO54 (SPegt-NheI/SacI-as)
5'-CCTGCTACTGCTCCCAGCAGCAGTGAAAGAGTCCAAAGTGGCAGCATG-3'
SO55 (SPegt-NheI/SacI-s)
5'-CTAGCATGCTGCCACTTTGGACTCTTTCACTGCTGCTGGGAGCAGTAGCAGGAGCT-3'

Fig 5